1971

Classification Error of Gaussian and Transformed Gaussian Variates

John E. Boyd

CLASSIFICATION ERROR OF GAUSSIAN AND TRANSFORMED GAUSSIAN VARIATES

BY

JOHN E. BOYD

A thesis submitted
in partial fulfillment of the requirements for the
degree Master of Science, Major in
Electrical Engineering
South Dakota State University

1971

CLASSIFICATION ERROR OF GAUSSIAN AND
TRANSFORMED GAUSSIAN VARIATES

This thesis is approved as a creditable and independent investigation by a candidate for the degree, Master of Science, and is acceptable as meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Thesis Advisor Date

Head, Electrical Engineering Date

ACKNOWLEDGMENT

The author expresses his appreciation to Dr. Gerald Nelson, who provided invaluable assistance in the interpretation and comprehension of fundamental decision theory and of the research results. Others on the Electrical Engineering staff were unselfish in giving of their time and ideas. The author extends a sincere thank you to Dr. Frank Fitchen and the National Science Foundation for the funding which enabled this research to be performed.

TABLE OF CONTENTS

CHAPTER 1: DEFINITION OF RESEARCH OBJECTIVE AND REVIEW OF REFERENCED LITERATURE

I. The Communication Process ---------------------- 1

II. Definition of Research Objective --------------- 2

III. Preview of Thesis' Contents -------------------- 8

IV. Review of Referenced Literature ---------------- 11

CHAPTER 2: BASIC CONCEPTS REVIEWED

I. Review of Fundamental Signal Detection Theory -- 14

II. Decision Boundary Expressions for Commonly Encountered Probability Density Functions ------ 20

III. Representation of the Bivariate Gaussian Probability Density Function ------------------- 22

CHAPTER 3: COMPUTATION OF ALPHA AND BETA, RECEIVER OPTIMAL FOR UNIVARIATE CAUCHY DISTRIBUTION

I. Classification of Cauchy Distributed Random Variates -- 25

II. Calculation of Alpha and Beta ------------------ 29

III. Effect of Correlation Coefficient on Gaussian Distribution -------------------------------------- 31

CHAPTER 4: COMPUTATION OF ALPHA AND BETA, RECEIVER OPTIMAL FOR BIVARIATE GAUSSIAN DISTRIBUTION

I. Rotational Transformation for Uncorrelation of Correlated Variates ---------------------------------- 38

II. Transformation of the Contours of Constant Probability Density and of the Decision Boundary -- 42

III. Determination of the X Domain Variances --------- 46

IV. Computation of Alpha and Beta ------------------- 51

V. Volume Under the Probability Density Function --- 56

CHAPTER 5: SUMMARY OF OTHER INVESTIGATIVE STUDIES

I. Classification of Gaussian and Cauchy Variates by a Bayes' Receiver Optimal for a Gaussian Distribution ------------------------------------- 62

II. Transformation of Random Variates --------------- 68

III. Discussion of Transformation $z=x^2$ ---------------- 74

IV. The Weibull Distribution ------------------------ 77

CHAPTER 6: CONCLUSIONS ---------------------------------- 80

REFERENCES -- 86

APPENDIX 1: COMPUTATION OF THE ERROR FUNCTION -------- 88

APPENDIX 2: DERIVATION OF CAUCHY DISTRIBUTION FROM BIVARIATE GAUSSIAN DISTRIBUTION ---------- 91

APPENDIX 3: LISTING OF COMPUTER PROGRAMS -------------- 93

LIST OF FIGURES

Figure 1.1. Three-dimensional view of three probability density functions, characterized by $\sigma_{\alpha_1} = 1$, $\sigma_{\alpha_2} = 1$, and ρ -------------------- 4

Figure 1.2. Contours of two overlapping bivariate Gaussian probability density functions ----- 5

Figure 1.3. Two overlapping univariate Cauchy probability density functions -------------- 6

Figure 2.1. Decision boundary for Bayes' classification of two Gaussian distributions -------------- 19

Figure 3.1. Decision boundary for Bayes' classification of two univariate Cauchy distributions; $\mu_1 = 0$, $\mu_2 = 0.90$ -------------------------- 27

Figure 3.2. Decision boundary for Bayes' classification of two bivariate Gaussian distributions, $\rho = 0.98$ ----------------------------------- 32

Figure 3.3. Decision boundary for Bayes' classification of two univariate Cauchy distributions; $\mu_1 = 0$, $\mu_2 = 0.98$ -------------------------- 33

Figure 3.4. Two decision boundaries for two anti-correlated Gaussian distributions; $\rho_1 = 0.98$, $\rho_2 = -0.98$ ---------------------- 36

Figure 3.5. Decision boundary for two univariate Cauchy distributions; $\mu_1 = .98$, $\mu_2 = -.98$, $b_1 = 0.2$, $b_2 = 0.2$ ------------------------- 37

Figure 4.1. Definition of angle θ which describes a rotation transformation of two variates ---- 40

Figure 4.2. One-sigma contour of Gaussian probability density function for correlated and uncorrelated variates ---------------------- 50

Figure 4.3. Two-class bivariate Gaussian distribution after rotational transformations of variates ----------------------------------- 55

Figure 4.4. Volume contained under the Gaussian probability density function, within the 1, 2, and 3σ limits ---------------------- 58

Figure 4.5. Exact and approximate boundary for volume integration ------------------------ 60

Figure 5.1. Illustration of similarity between Gaussian and Cauchy probability density functions --------------------------------- 63

Figure 5.2. Block diagram of Cauchy-Gauss classification problem -------------------- 69

Figure 5.3. Three-sigma contours of constant probability density for two classes of distributions ---------------------------- 70

Figure 5.4. χ^2 probability density function for selected values of v --------------------- 76

Figure 5.5. Weibull probability density function for selected values of shape factor ----------- 78

LIST OF TABLES

Table 2.1. Decision boundary equations for 4 univariate and bivariate distributions ---------------------------------- 21

Table 5.1. Confusion matrix -- a priori Gaussian PDF Bayes' receiver ------------------------- 71

Table 6.1. Classification error for original and transformed distributions ------------------- 81

CHAPTER 1: DEFINITION OF RESEARCH OBJECTIVE AND REVIEW OF REFERENCED LITERATURE

I. The Communication Process

Communication is an act of making known to others information which one presently possesses. Whether between two individuals or ten nations, communication is the express-way for the exchange of opinions and desires. A process not limited to homo sapiens, communication can be an art form practiced in as many ways as there are participants. Effective communication requires that one correctly receive the information which was sent; this statement is as applicable to the transmission and reception of accounts from one bank to another as it is to the interpretation of diplomatic statements exchanged between heads of government. The task of making the correct decision is simplified if the decision to be made is binary; that is, if the decision is to be either yes or no. But even this relatively simple problem can pose difficult questions if considered in an engineering context. The techniques of making the correct decision form the core of this thesis; the significance of the message and the consequences of making the wrong decision are the important considerations from a communications viewpoint.

Pattern recognition or signal detection is the broad heading under which a study such as this thesis is generally categorized. The binary signals to be classified, which are termed either class one or class two, have no message significance and in fact are known a priori, that is, beforehand. The consequence of making a wrong decision is that each misclassification will be recorded, along with the correct decisions, in a performance index called the confusion matrix. If the level of performance of a receiver (also called a classifier) is being predicted, rather than being measured, the fraction of incorrect decisions the receiver will make over the long run is termed the classification error. Two types of classification error are defined in Chapter 2.

II. Definition of Research Objective

As indicated by the title and by the large proportion of pages devoted to the subject, transformation of variables is the primary topic of this thesis. The transformation of partially overlapping, correlated, two-dimensional, Gaussian distributed random variates to partially overlapping, independent, one-dimensional, Cauchy distributed random variates involves a reduction in dimensionality by one. Consequently, the equation defining the receiver is simplified, and the expressions for predicting the classification error are easier to formulate. The equation required to effect this transformation is $z=Y_1/Y_2$, where Y_1,Y_2 are the standardized correlated Gaussian variates and z is the independent Cauchy random variate.

To aid in understanding the concept that the probability density function actually does change as a result of the transformation, observe Figure 1.1-Figure 1.3. The 3-dimensional views in Figure 1.1 illustrate the probability density functions of three bivariate Gaussian distributions having variances equal to unity and correlation coefficient equal to ρ. Figure 1.1(a) shows the class one distribution of independent variates, and Figure 1.1(c) illustrates the class two distribution. The circular or ellipsoidal boundaries indicate contours of constant probability density at distances of 1σ, 2σ, and 3σ from the mean. (Standard notation defines σ as the standard deviation, μ as the mean, and ρ as the correlation coefficient.)

An overhead view of Figure 1.1(a) and 1.1(c), with the means of the class one and class two distributions located at coincident points, appears in Figure 1.2. The contours of constant probability density are again visible, but this time their usefulness may be more readily recognized. A narrowing and lengthening of the class two probability density function illustrates the considerable effect of the correlation between variates. The declination of the major axis also varies with ρ, and is 45° in Figure 1.2; in Chapter 4 an equation is given which defines the declination angle, based upon the correlation coefficient and variances.

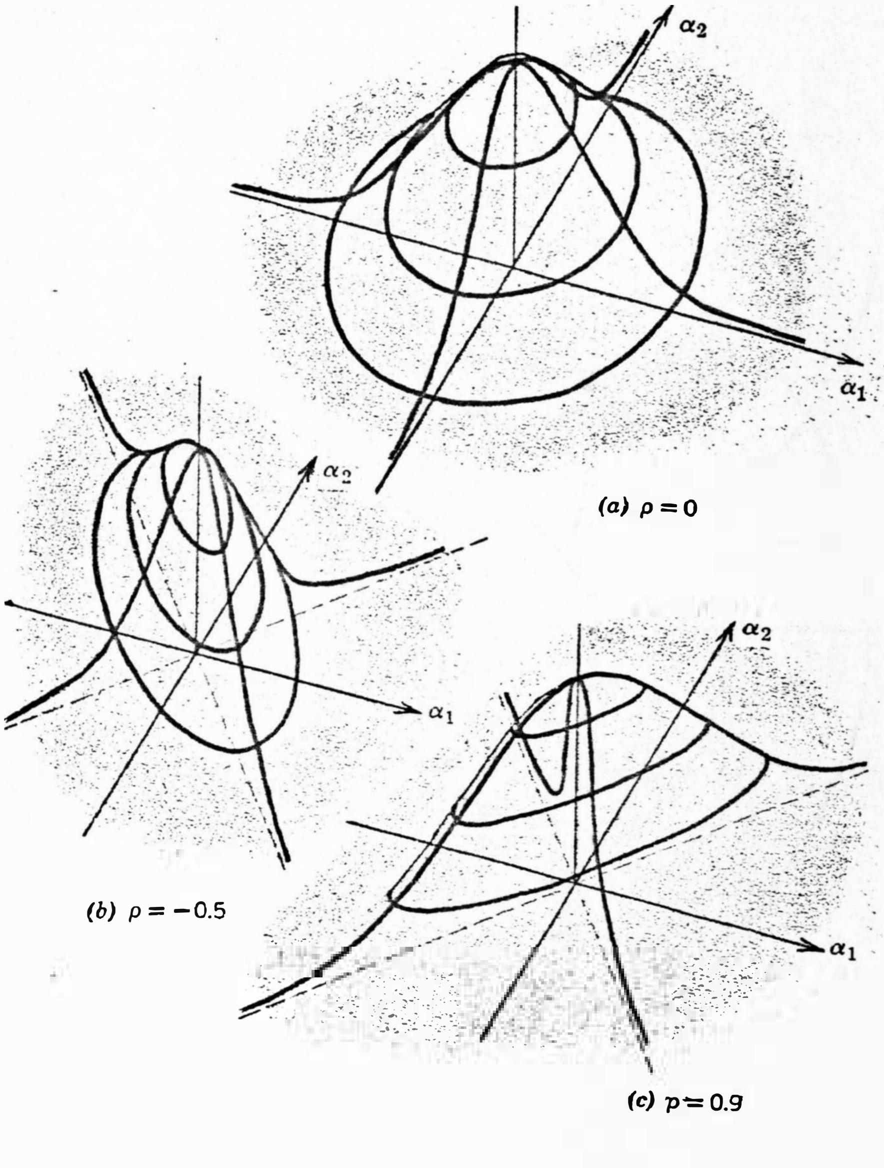


Figure 1.1. Three-dimensional view of 3 probability density functions, each characterized by $\sigma_{\alpha_1}=1$, $\sigma_{\alpha_2}=1$, and ρ.

Reproduced with permission from John Wiley and Sons, New York. Wozencraft and Jacobs, [2], p. 52.

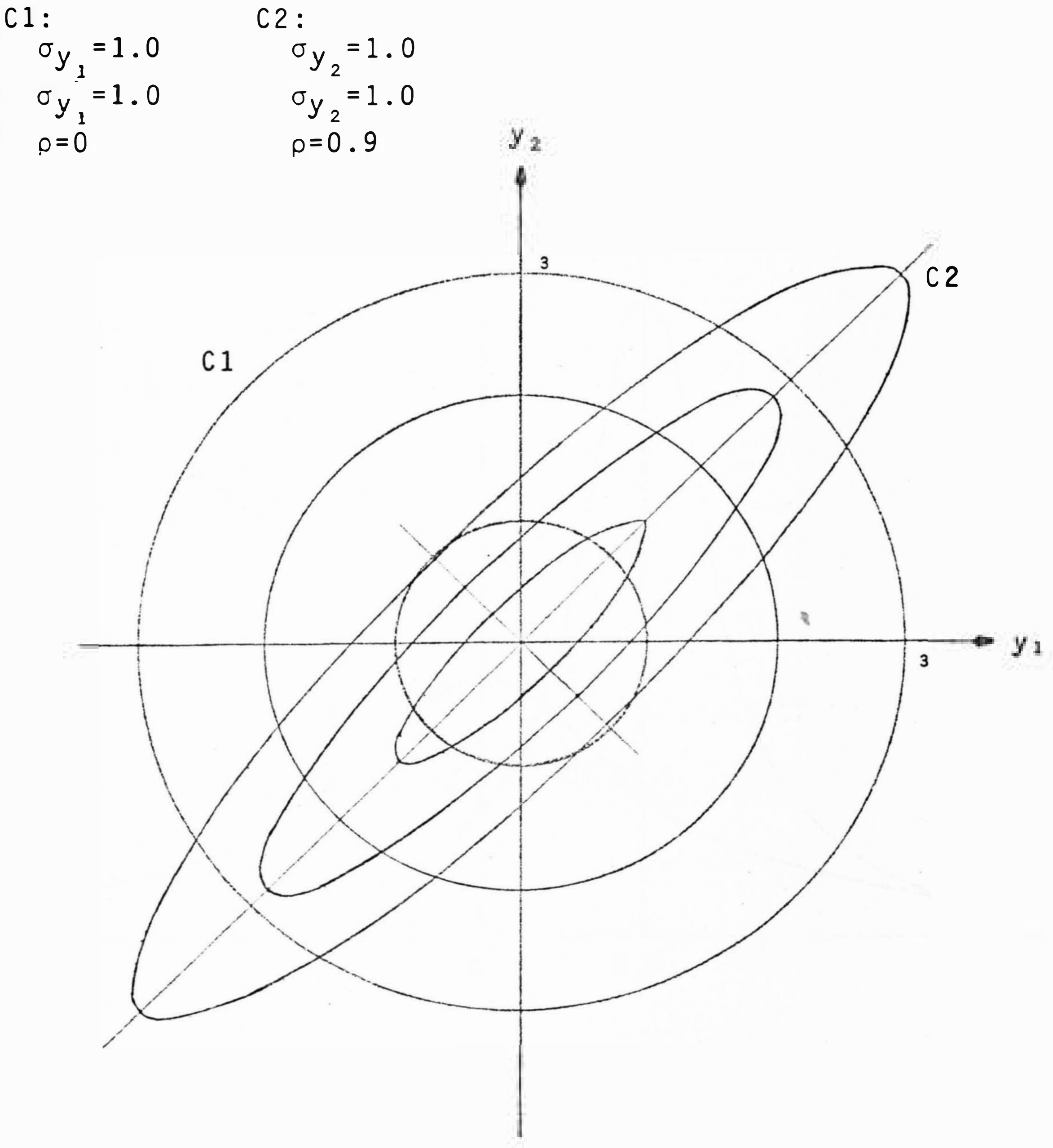


Figure 1.2. Contours of two overlapping bivariate Gaussian probability density functions.

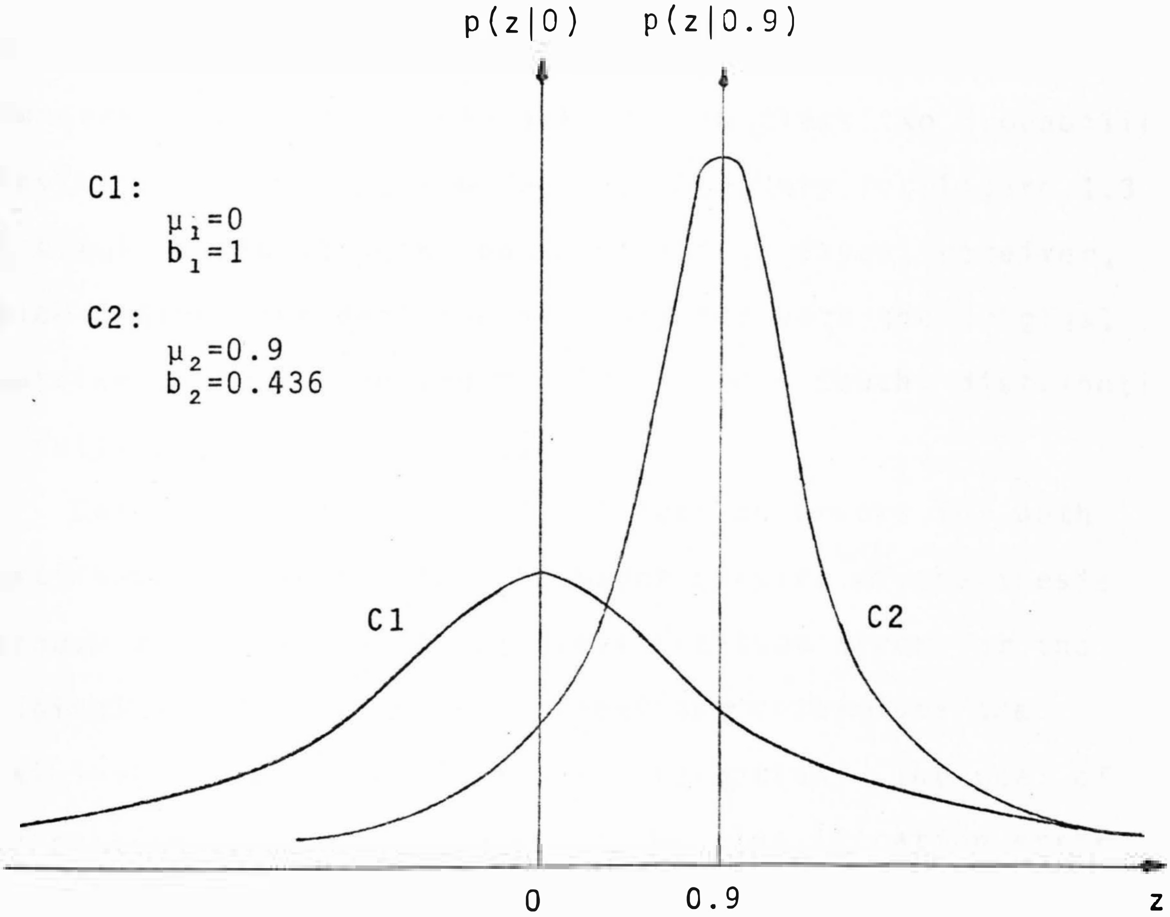


Figure 1.3. Two overlapping univariate Cauchy probability density functions.

In Figure 1.3 the probability density function of the transformed random variate z illustrates the reduction in dimensionality. Clearly, the two Cauchy distributions, represented by the two probability density functions, are univariate. As will be shown in Chapter 3 and 4, the decision boundary for Figure 1.2 consists of two hyperbolas symmetrical with the minor axis of the class two probability density function, but the decision boundary for Figure 1.3 is simply a single point on the z axis. Bayes' receiver, which defines the decision boundary for both the original Gaussian distribution and the transformed Cauchy distribution, is fully optimum in each case.

Determination of the classification errors for both distributions constitutes the major portion of the thesis, because a comparison of the classification error for the original and transformed distributions determines the usefulness (goodness) of the transformation. The goal of the research was to determine if the classification error of the Cauchy distribution was as low, or lower, than the classification error of the Bayes' receiver optimum for the bivariate Gaussian distribution. If lower, the transformation achieves a reduction in error, an ever desirable achievement; if the error remains approximately constant, the transformation is useful as a means for approximating the classification error because the classification error for the Cauchy distribution is substantially easier to formulate than that for the Gaussian distribution.

A secondary objective was to evaluate other transformations, particularly the familiar $y=x^2$ transformation, if a decision boundary could be formulated. Further, an estimate of the validity of approximating the Gaussian distribution by a Cauchy distribution was sought.

III. Preview of Thesis' Contents

Because several topics of research effort are discussed in the thesis, a brief commentary on these topics may aid the reader in following the development of succeeding chapters. Chapter 3 discusses the decision boundary required to classify Cauchy random variates, defines the classification error for that receiver, and shows what effect the correlation coefficient has upon the mean of the Cauchy distribution. Two sets of classification error are computed, one for a correlation coefficient of the Gaussian random variates of 0.90, and one for 0.98.

In Chapter 4 the decision boundary required to classify a priori Gaussian variates and to compute the classification error is derived. However, in order to proceed with the derivation, a rotational transformation capable of changing correlated variates to uncorrelated is required. A summary of the rotational transformation theory, and its application to the decision boundary (receiver) equation and probability density function, forms the first half of the chapter. Evaluation of classification error, and verification of the values obtained constitute the remainder of the chapter.

Although the word receiver was inserted parenthetically after the term decision boundary, the two are not precisely equivalent. The distinction between them is made in Chapter 2. Also discussed is a summary of basic signal detection theory. Because it has not been found elsewhere, a table defining the decision boundary for both univariate and bivariate probability density functions of the Gauss, Cauchy, Rayleigh, and Weibull distributions is included in Chapter 2.

Other areas of investigation which were not researched as vigorously as was the transformation $z=Y_1/Y_2$ are combined to form Chapter 5. The first topic, the classification of Cauchy variates by a receiver designed for a priori Gaussian variates, is discussed. Three different sets of overlapping distributions are classified. Although the simulation provides an indication of the merit of the study, the potential for further investigation and additional simulation is by no means exhausted. Tabular results of those sets classified are included. A brief review of multivariate transformations is discussed next; although helpful to the author's understanding of transformations, no new results are reported.

In section three of Chapter 5, the transformation $y=x^2$ is applied to a univariate Gaussian density function. The class one distribution, centered at zero, transforms to a chi-squared distribution; the class two distribution of mean μ transforms to a non-central chi-squared distribution. The complexity of the probability density function of the latter distribution prevents further study. Although a receiver can be formulated, expressions for classification error prove to be unattainable by analytical methods. The fourth study reported in Chapter 5 is formulation of a receiver designed for the Weibull distribution. A Weibull probability density function is equal to the exponential and Rayleigh, and approximates the Gaussian, probability density function for certain values of the shape factor s. Once again, it is found that classification error cannot be computed by analytical methods.

Chapter 6 briefly comments on the results of the research discussed in the previous five chapters. Subjects already pursued which appear to offer further opportunities for research are mentioned in the hope that others may benefit from the preliminary investigation.

A discussion of the difficulties encountered when calculating the error function with a digital computer forms Appendix 1. Comments pertain to the limitations of the computer when called upon to handle large numbers. The computer programs used in the evaluation of the classification error in Chapter 4 and the simulation study of Chapter 5 are contained in Appendix 3. In addition, a derivation of the Cauchy distribution from the Gaussian distribution is contained in Appendix 2.

IV. Review of Referenced Literature

Only a small fraction of the several score of books available on the subject of statistics and pattern recognition was used as reference material. A brief summary of the books providing information used in the thesis is included in the belief that the reader may wish to refer directly to the original source. Hancock and Wintz [1] is the main reference for the fundamental concepts embodied in this thesis, particularly in Chapter 2. Definitions of alpha and beta error, optimum Bayes' receiver, and observation space contained in their book are basic. Another text frequently referenced is Wozencraft and Jacobs [2]. Many basic concepts of one and two-dimensional random variables, as well as elementary transformations of variables are presented. A good treatment of the Gaussian process, especially the bivariate distribution, is found in Davenport and Root [3]. Papoulis [4] provided the theory that prompted the research forming the main topic of this thesis; of particular interest is his lucid interpretation of one function of two random variates. An algorithm for calculating recognition error when applying data vectors from two Gaussian populations to an optimum Bayes' classifier is presented by Fukunaga and Krile [5]. Their method appears to provide good results when classifying up to eight-dimensional Gaussian data having unequal covariance matrices and arbitrary a priori probabilities. Their algorithm is a general solution to a problem analogous to that discussed in this thesis.

Anderson [6] briefly discusses transformation of variables, and covers in greater detail a derivation of the non-central chi-squared distribution. Hogg and Craig [7] develop the theory on Gaussian and chi-squared distributions; they also discuss transformations of several variables. In Johnson and Leone [8] one finds an abbreviated disquisition of continuous distributions, including the normal, exponential, chi-squared, and Weibull; in addition, they devote a section to transformation of variables, and one to the derivation of the chi-squared distribution. Techniques of computer generation of discrete and continuous probability distributions are discussed in Naylor [9].

Hildebrand [10] includes methods of performing numerical integration; he discusses the trapezoidal rule and Simpson's rule, as well as the parabolic and Newton-Cotes formulas. Abramowitz and Stegun [11] provide equations for computation of error function, and tabulate the error function for arguments up to 2.0. Although not rigorously developed, Blachman [12] presents a readable discussion on the univariate, bivariate, and multivariate Gaussian distribution. And, Wee's [13] bibliography includes ninety-six references listed by categories on the broad topic of pattern recognition.

CHAPTER 2: BASIC CONCEPTS REVIEWED

I. Review of Fundamental Signal Detection Theory

The author assumes the reader is acquainted with probability theory, random variables, and statistics; accordingly, the theory to be developed in this chapter will focus on elementary concepts of signal detection. The content of this chapter is based largely on Chapter 3 of Hancock and Wintz [14].

Consider the random variable described by

$$p(x_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x_1-\mu)^2}{2\sigma^2}\right] \quad \text{or} \tag{1}$$

$$p(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{\frac{(x_1-\mu_1)^2\sigma_2{}^2-2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)+(x_2-\mu_2)^2\sigma_1{}^2}{\sigma_1{}^2\sigma_2{}^2}\right\}\right]. \tag{2}$$

Define the parameter vector $\underline{\theta} = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$ or $\underline{\theta} = \begin{bmatrix} \mu \\ \sigma \\ \rho \end{bmatrix}$,

and the data vector $\underline{X} = [x_1]$ or $\underline{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

If all of the parameters are uniquely determined, then $\underline{\theta}$ is called a simple hypothesis. If, for example, $0<\mu<0.5$, $\sigma=1$, $\rho=0.9$, then $\underline{\theta}$ is termed a composite hypothesis. In this thesis μ, σ, and ρ are uniquely known; therefore the following procedure is known as simple hypothesis testing.

For this thesis only two values of $\underline{\theta}$ are possible, $\underline{\theta}=\underline{\theta}_1$, or $\underline{\theta}=\underline{\theta}_2$. The former may be termed the null hypothesis H_0, and the latter termed the alternate hypothesis H_1. One can either accept H_0, or reject it. The test procedure for making the decision is called the decision rule, generally formulated by partitioning the observation space Γ into two disjoint subspaces Γ_1 and Γ_2. One decides in favor of H_0 if the data vector is contained in Γ_1, and rejects H_0 if the data vector is contained in Γ_2. This result indicates it is possible to commit two kinds of errors, called alpha and beta.

Alpha is the probability of rejecting H_0 when it is true; alpha sometimes is termed error of the first kind.

$$\alpha = \int_{\Gamma_2} p(\underline{X} \mid \underline{\theta}_1) d\underline{X} \tag{3}$$

Beta is defined as the probability of accepting H_0 when it is false; beta sometimes is termed error of the second kind.

$$\beta = \int_{\Gamma_1} p(\underline{X} \mid \underline{\theta}_2) d\underline{X} \tag{4}$$

Consequently, $(1-\alpha)$ is defined as the probability of making a correct decision when H_0 is true, and $(1-\beta)$ is defined as the probability of making a correct decision when H_1 is true.

For the problem discussed in this thesis one may assign a priori probabilities to the events defined by $\underline{\theta} = \underline{\theta}_1$ and $\underline{\theta} = \underline{\theta}_2$ as being q_1 and q_2, respectively. It is desirable, in the general solution at least, to include cost functions which describe the consequences of making a correct decision and committing an error in decision. C_{ij} is the cost to the observer if he decides hypothesis H_i is true when H_j is true; $i, j = 0,1$. Average risk R can now be defined.

$$R = q_1 C_{00} + q_2 C_{11} + q_1 (C_{10} - C_{00}) \alpha + q_2 (C_{01} - C_{11}) \beta$$

For most problems let the cost of making a correct decision, C_{00} and C_{11}, be zero. Define C_{10} as being equal to C_α, and C_{01} as being equal to C_β.

$$R = \alpha q_1 C_\alpha + \beta q_2 C_\beta \tag{5}$$

Now two important quantities may be defined, the likelihood ratio $\Lambda(\underline{X})$ and the threshold K.

$$\Lambda(\underline{X}) = \frac{p(\underline{X} \mid \underline{\theta}_2)}{p(\underline{X} \mid \underline{\theta}_1)} \qquad K = \frac{q_1 C_\alpha}{q_2 C_\beta} \tag{6}$$

The decision boundary is determined by equating $\Lambda(\underline{X})=K$, which establishes the boundary defining the two subspaces Γ_1 and Γ_2 of the sample space Γ. A particular data vector $\underline{X}'$ may be classified by comparing the likelihood ratio to the threshold. If $\Lambda(\underline{X}')>K$, decide $\underline{X}'$ came from $p(\underline{X} \mid \underline{\theta}_2)$; i.e., decide in favor of the alternate hypothesis. Similarly, if $\Lambda(\underline{X}')<K$, decide $\underline{X}'$ came from $p(\underline{X} \mid \underline{\theta}_1)$; i.e., decide in favor of the null hypothesis.

Bayes' solution is the strategy of determining a decision boundary on the basis of the smallest average risk, called Bayes' risk. In later chapters of this thesis the author assumes $q_1=q_2$, and $C_\alpha=C_\beta$. Therefore, from (6) the threshold K is determined to be unity.

To illustrate the procedure for obtaining the decision boundary in general, and therefore the classifier, consider the simple example of two distributions of an independent Gaussian variate. H_0 has zero mean and variance σ^2, and H_1 has mean μ and variance σ^2. Equate the likelihood ratio to the threshold to obtain the decision boundary:

$$K = \Lambda(x_1) = \frac{p(x_1|\mu)}{p(x_1|o)} = \frac{1/\sqrt{2\pi}\ \sigma\ \exp\ [-(x_1-\mu)^2/2\sigma^2]}{1/\sqrt{2\pi}\ \sigma\ \exp\ [-x_1{}^2/2\sigma^2]} \tag{7}$$

After solving for x_1 one obtains

$$x_1 = \frac{\mu}{2} + \frac{\sigma^2}{\mu}\ \ln K. \tag{8}$$

Equation (8) is the decision boundary, and it can be shown to be intuitively correct if K is set equal to unity. Under this restraint ln K=0, and the second term of (8) goes to zero. The decision boundary expression reduces to

$$x = \frac{\mu}{2}.$$

Placing the decision boundary at the midpoint of the two means is an intuitively logical answer. Figure 2.1 illustrates the observation space and the two disjoint subspaces, as well as the decision boundary and classification errors alpha and beta.

For two classes of data not having equal variances, and where one mean is nonzero, the equation for the decision boundary is not obvious and may not be so easily checked. Refer to Table 2.1 for decision boundary equations of Gaussian, Rayleigh, Cauchy, and Weibull distributions.

$$\text{DB: } \frac{\mu}{2} + \frac{\sigma^2}{\mu} \ln K$$

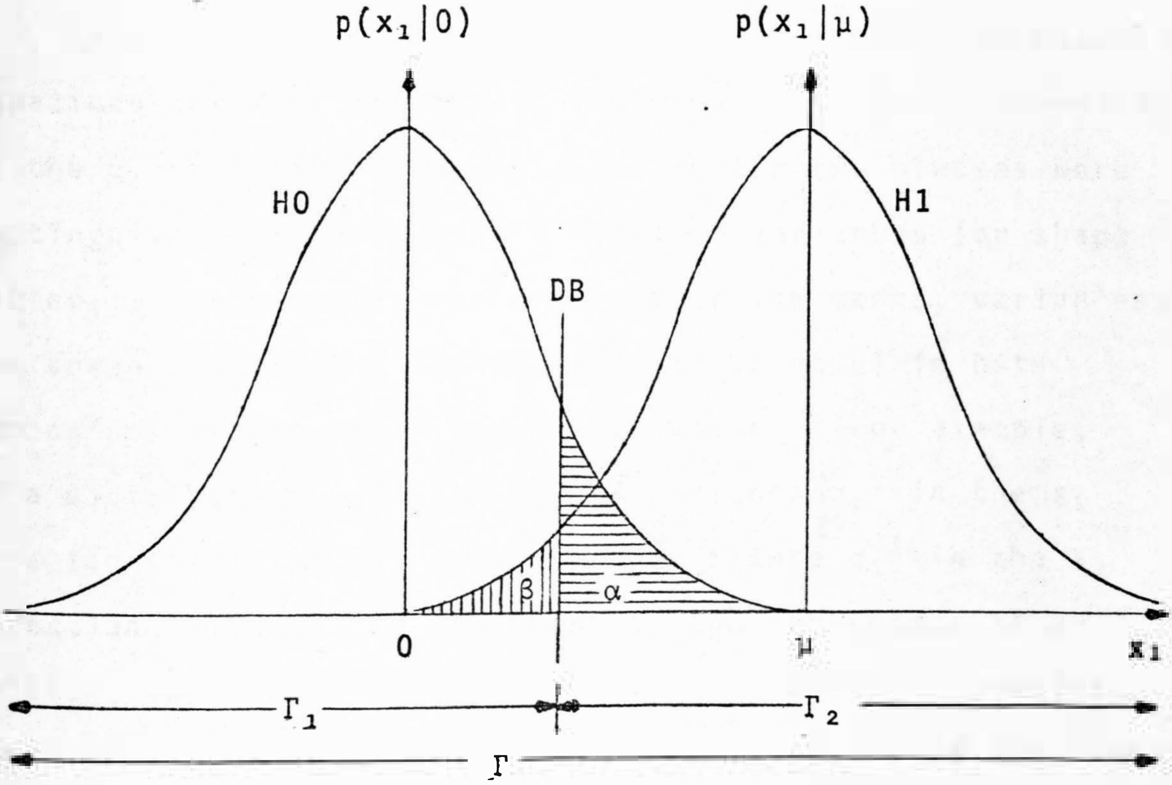


Figure 2.1. Decision boundary for Bayes' classification of two Gaussian distributions.

II. Decision Boundary Expressions For Commonly Encountered Probability Density Functions

A brief review of theory describing the formulation of a decision boundary based upon knowledge of the likelihood ratio $\Lambda(\underline{X})$ and threshold K was provided in the previous section. The decision boundary was shown to be a partition separating the observation space Γ into two subspaces Γ_1 and Γ_2.

Included in this section is a table of decision boundary equations for four continuous distributions. The parameters of the distributions were selected so the two classes were distinguished either by their means or variances (or shape factors). To simplify the expressions the means, variances, and shape factors are always taken to be equal in both dimensions of the bivariate distributions. For example, if a distribution has mean μ_1 and variance $\sigma_1{}^2$ in the x_1 direction, it also has mean μ_1 and variance $\sigma_1{}^2$ in the x_2 direction. This constraint permits the subscripts to be omitted, and the parameters written as μ and σ^2. However, subscripts are required to distinguish variance of the class one distribution from variance of the class two distribution; the parameter $\sigma_1{}^2$ is read as the variance of the class one distribution.

Only for the univariate Gaussian and Cauchy distributions is it possible to express the variable as an explicit function

Table 2.1. Decision boundary equations for 4 univariate and bivariate distributions

1	GAUSS	$x=-\frac{(\mu_1\sigma_2^2-\mu_2\sigma_1^2)}{\sigma_1^2\sigma_2^2}\pm\frac{\sigma_1^2\sigma_2^2}{(\sigma_1^2-\sigma_2^2)}\left(2\sigma_1^2\sigma_2^4\ln\left[\frac{K\sigma_1}{\sigma_2}\right]+\sigma_1^2\sigma_2^2(\mu_1-\mu_2)^2-2\sigma_1^4\sigma_2^2\ln\left[\frac{K\sigma_1}{\sigma_2}\right]\right)^{1/2}$
1	CAUCHY	$x=\frac{K\mu b_1\pm\sqrt{b_1b_2(\mu^2+b_1^2+b_2^2)K-b_1b_2(K^2+1)}}{(Kb_1-b_2)}$
1	RAYLEIGH	$\ln\left[\frac{x-\mu}{x}\right]+\frac{1}{2b_1^2b_2^2}\left[(b_2^2-b_1^2)x^2+(2\mu b_1^2)x-\mu^2b_1^2\right]=\ln\left[\frac{b_2^2K}{b_1^2}\right]$
1	WEIBULL	$(s-1)\ln\left[\frac{(x-l_2)}{(x-l_1)}\right]+(x-l_1)^s-(x-l_2)^s=\ln[K]$ for $s_1=s_2=s$, $(s_2-s_1)\ln[x]+x^{s_1}-x^{s_2}=\ln\left[\frac{Ks_2}{s_1}\right]$ and for $l_1=l_2=l$
2	GAUSS	$2\sigma_1^2\sigma_2^2\ln\left[\frac{\sigma_2^2K}{\sigma_1^2}\right]+2\mu^2\sigma_1^2=(\sigma_2^2-\sigma_1^2)x_1^2+(\sigma_2^2-\sigma_1^2)x_2^2+2\mu\sigma_1^2(x_1+x_2)$
2	CAUCHY	$\left[(Kb_1^2b_2^2+Kb_1^2\mu^2-2b_1^2b_2^2)x_1^2+(Kb_1^2b_2^2+Kb_1^2\mu^2)x_2^2+(Kb_1^2-b_2^2)x_1^2x_2^2-(2Kb_1^2\mu)x_1^2x_2-(2Kb_1^2\mu)x_1x_2^2-(2Kb_1^2b_2^2\mu+2Kb_1^2\mu^3)x_1^2-(2Kb_1^2b_2^2\mu+2Kb_1^2\mu^3)x+(4Kb_1^2\mu^2)x_1x_2\right]=b_1^4b_2^2-Kb_1^2(2\mu^2b_2^2+b_2^4+\mu^4)$
2	RAYLEIGH	$\ln\left[\frac{x_1-\mu}{x_1}\right]+\ln\left[\frac{x_2-\mu}{x_2}\right]+\frac{x_1^2+x_2^2}{2b_1^2}-\frac{(x_1-\mu)^2+(x_2-\mu)^2}{2b_2^2}=\ln\left[\frac{b_2^4K}{b_1^4}\right]$
2	WEIBULL	$(s+1)\ln\left[\frac{x_2-l_2}{x_2-l_1}\right]+(s-1)\left(\frac{x_1-l_2}{x_1-l_1}\right)-(x_2-l_2)^s+(x_2-l_1)^s-(x_1-l_2)^s+(x_1-l_1)^s=\ln[K]$

of the parameters; in all other cases the equation is transcendental. Variates may still be classified by a Bayes' receiver, but alpha and beta cannot easily be computed because the decision boundary equation is not known as a function of x, but rather as a function of ln $(x-\mu)$.

III. Representation of the Bivariate Gaussian Probability Density Function

Although mathematically one can describe a k-dimensional observation space, he soon reaches the limits of graphical representation. The probability density function of a one-dimensional data vector $[x_1]$ is easy to represent by a two-dimensional graph, such as that in Figure 2.1. However, the three-dimensional figure required for a two-dimensional data vector $[x_1 x_2]$ is much more difficult to draw, and depending on the manner by which it is depicted, may be difficult to comprehend as well. Figures 1.1(a) and 1.2 illustrate two methods of representing a three-dimensional probability density function. The first is easier to comprehend, but the second supplies more factual information.

From a drawing such as Figure 1.2 one should note that the variances of the class two probability density function are not as specified, but rather are modified by the correlation coefficient. The concentric ellipses cross the y_1, y_2 axis at points given by the equation

y_1 intercept $= \sqrt{1-\rho^2}\,\sigma_{y_1}$, y_2 intercept $= \sqrt{1-\rho^2}\,\sigma_{y_2}$.

Blachman [15] terms the circular or elliptic boundaries, contours.

The contours of constant probability density in the $y_1 y_2$ plane are the loci defined by the exponent of (2), where this exponent is equal to a constant. The required equations for the contours of constant probability density, written for correlated and uncorrelated variates, are

$$(1-\rho^2)\sigma^2 = \left(\frac{y_1^{\,2}}{\sigma_{y_1}^2} - \frac{2\rho y_1 y_2}{\sigma_{y_1}\sigma_{y_2}} + \frac{y_2^{\,2}}{\sigma_{y_2}^2} \right) \tag{10}$$

$$\sigma^2 = \left(\frac{x_1^{\,2}}{\sigma_{x_1}^2} + \frac{x_2^{\,2}}{\sigma_{x_2}^2} \right). \tag{11}$$

Variance σ^2 in the above equations defines which contour is being described; σ^2 typically is 1, 2, or 3, but may be any desired fraction.

One may question the author's purpose in writing the uncorrelated variates in a new domain, rather than referring to the x_1, x_2 variates in terms of the Y domain variables. In Chapter 4 the motivation will become apparent; a transformation of correlated variates to uncorrelated variates by means of a rotational transformation will be described. The transformation permits considerable simplification in the equations used to describe the contours of constant probability density and to draw the decision boundary. Two domains, X and Y, are required to describe the appropriate variates. It is also for this reason that the variances are subscripted by x or y where any opportunity for ambiguity exists.

CHAPTER 3: COMPUTATION OF ALPHA AND BETA, RECEIVER OPTIMAL FOR UNIVARIATE CAUCHY DISTRIBUTION

I. Classification of Cauchy Distributed Random Variates

Numerical values for the decision boundary, alpha error and beta error are calculated in this chapter, which are compared to the corresponding values computed for the two class bivariate Gaussian distributions in Chapter 4.

The univariate Cauchy probability density function is defined by the equation

$$p(x_1) = \left(\frac{b/\pi}{(x_1-\mu)^2+b^2}\right). \tag{1}$$

A derivation of (1) from two jointly Gaussian random variates, using the transformation $z = y_1/y_2$, is contained in Appendix 2. The important result of the derivation is written below.

$$p(z) = \left(\frac{\sqrt{1-\rho^2}\quad \sigma_{y_1}\sigma_{y_2}/\pi}{\sigma_{y_2}^2\quad (z-\rho\sigma_{y_1}/\sigma_{y_2})^2+\sigma_{y_1}^2\quad (1-\rho^2)}\right) \tag{2}$$

Comparing (2) with (1), variables b and μ are defined in terms of σ_y and ρ.

$$b = \frac{\sqrt{1-\rho^2}\ \sigma_1}{\sigma_2} \qquad \mu = \frac{\rho\sigma_1}{\sigma_2} \tag{3}$$

An interesting result is contained in (2); a correlated, bivariate Gaussian distribution transforms to an uncorrelated, univariate Cauchy distribution centered at $\mu = \rho\sigma_1/\sigma_2$. The observation space has been decreased by one dimension, and the mean of the transformed distribution is proportional to the correlation coefficient of the original distribution. The correlation coefficient of the Gaussian random variates may be zero, in which case the Cauchy distribution is centered at zero and the shape of the probability density function is determined by $b = \sigma_1/\sigma_2$. Negatively correlated variates cause the Cauchy distribution to be shifted left of zero. In this thesis the main investigation is based on the fact that one class of variates is capable of being shifted a distance $\rho\sigma_1/\sigma_2$ left or right of zero. The simulation discussed in Chapter 5 also depends on this fact.

As explained in Chapter 1, the primary purpose of this research was to determine if the transformed Gaussian distributed random variate could be classified with lower error than the original Gaussian random variates. One important point is repeated for emphasis. The Bayes' receiver defined for the univariate Cauchy distribution is optimum, just as the Bayes' receiver defined for the bivariate Gaussian distribution is optimum. No other classification procedure will result in a lower value of alpha and beta for the respective distributions.

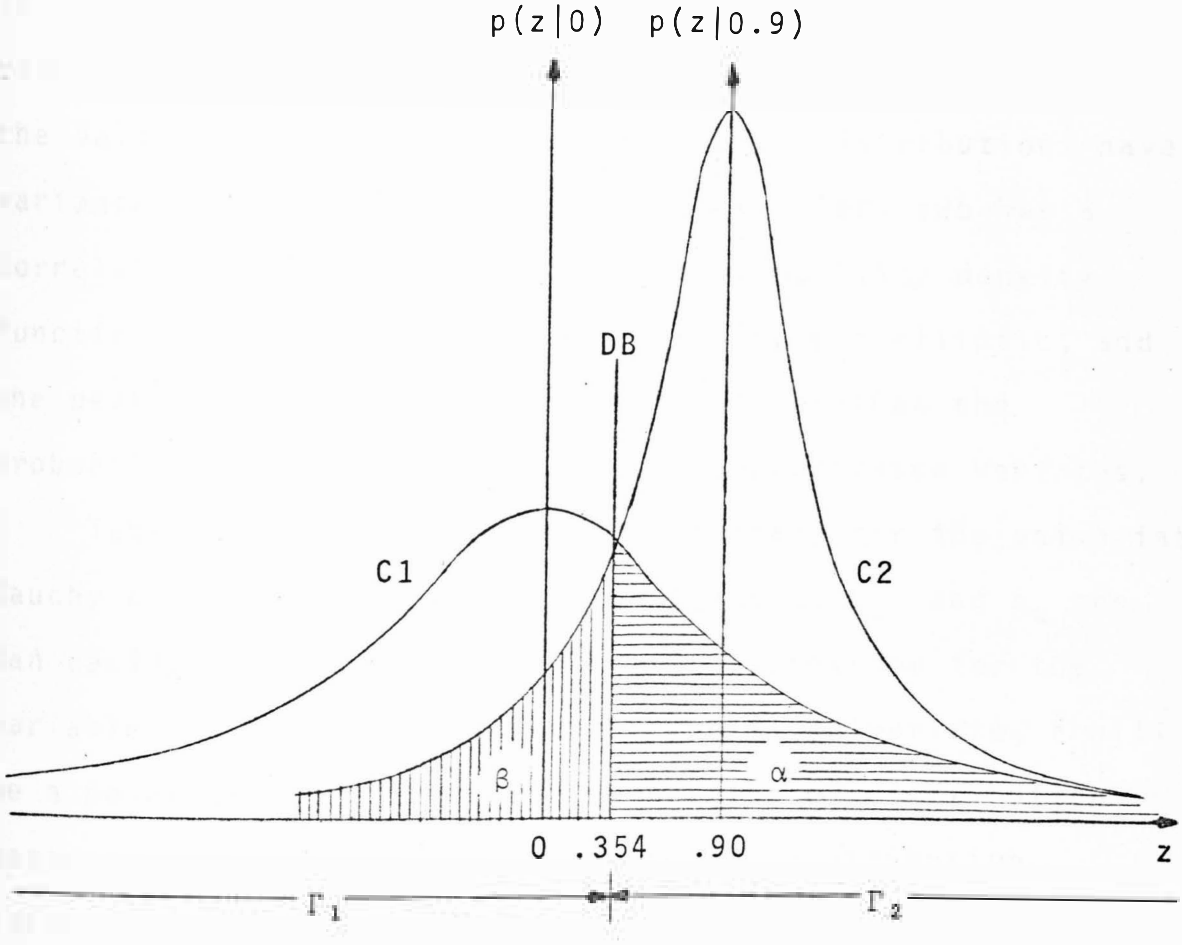


Figure 3.1. Decision boundary for Bayes' classification of two univariate Cauchy distributions; $\mu_1=0$, $\mu_2=0.90$.

Proceeding toward the goal of solving the problem, within the self-imposed constraint of considering only two numerical examples, the first step is to evaluate alpha and beta for the two-class Cauchy distribution. To refresh the reader's familiarity of the parameters used, the following information is repeated. Both Gaussian distributions have mean (0,0), a requirement which must be observed to maintain the validity of the transformation; both distributions have variances each equal to unity. Because class two has a correlation coefficient of 0.9, its probability density function contours are no longer circular but elliptic, and the peak of the density function is higher than the probability density function of the uncorrelated variates.

Table 2.1 lists the decision boundary for the univariate Cauchy distribution. After determining μ, b_1, and b_2 one can easily solve the given quadratic expression for the variable z. Because the distribution is univariate, z will be a point on the z axis, not a function of distribution parameters. The equation for z and the distribution parameters are repeated below for convenience.

$$Z = \left\{ \frac{K\mu b_1 \pm \sqrt{b_1 b_2 [(\mu^2 + b_1{}^2 + b_2{}^2)\ K - b_1 b_2 (K^2+1)]}}{K b_1 - b_2} \right\} \tag{4}$$

$$\mu = \frac{\rho\sigma_1}{\sigma_2} = 0.9 \qquad b_1 = \frac{\sigma_1}{\sigma_2} = 1.0 \qquad b_2 = \frac{\sqrt{1-\rho^2}\,\sigma_1}{\sigma_2} = 0.436 \qquad K = 1.0$$

Substituting the given values into (4) and solving, z=0.354, 2.84. Referring to Figure 3.1 the correct answer is obviously 0.354. Data vectors $[z_1]$ occurring to the right of z=0.354 are classified as class two, and those to the left as class one. To determine the percentage of errors resulting from this optimum selection of the decision boundary, one must compute alpha and beta.

II. Calculation of Alpha and Beta

In Chapter 2, (3) and (4) define the alpha and beta classification error in terms of the probability density functions of the two classes and the observation subspaces Γ_1 and Γ_2. In (3), alpha is defined as

$$\alpha = \int_{\Gamma_2} p(\underline{X}|\underline{\theta}_1)d\underline{X}. \tag{5}$$

Substitution of the Cauchy probability density function into (5) yields

$$\alpha = \frac{\sigma_1/\sigma_2}{\pi} \int_{DB}^{\infty} \frac{dz}{z^2+(\sigma_1/\sigma_2)^2},$$

where DB stands for the decision boundary point. The known values are substituted for σ_1, σ_2, giving

$$\alpha = \frac{1}{\pi} \int_{0.354}^{\infty} \frac{dz}{z^2+1},$$

which integrates as the arctan of z. Alpha equals 0.392.

Calculation of beta proceeds in a similar manner. Equation (4) defines beta as

$$\beta = \int_{\Gamma_1} p(\underline{X}|\underline{\theta}_2)d\underline{X} \ . \tag{6}$$

Substitution of the Cauchy probability density function into (6) yields

$$\beta = \int_{-\infty}^{DB} \left(\frac{\sqrt{1-\rho^2}\ \sigma_1\sigma_2}{\pi}\right) \frac{dz}{\sigma_2{}^2(z-\rho\sigma_1/\sigma_2)^2+\sigma_1{}^2(1-\rho^2)} \ .$$

Known values are substituted, and a change of variables $x = (z-0.9)$ made which results in

$$\beta = \frac{0.436}{\pi} \int_{-\infty}^{-0.546} \frac{dx}{x^2+(0.436)^2} \ .$$

Beta integrates as the arctan of x, giving a final answer of beta equal to 0.216.

Although not immediately obvious from the computations, the time required to compute alpha and beta for the Cauchy optimal Bayes' receiver is miniscule in comparison to similar computations required for alpha and beta of the Bayes' receiver optimal for the bivariate correlated Gaussian distribution presented in the next chapter.

III. Effect of Correlation Coefficient on Gaussian Distribution

The effect which the correlation coefficient ρ has upon the probability density function of the original bivariate Gaussian distribution and the transformed Gaussian distribution can be ascertained by comparing Figures 1.2 and 1.3 with Figures 3.2 and 3.3. In the latter figures the correlation coefficient ρ has increased to 0.98. A further narrowing and lengthening of the class two Gaussian distribution has occurred, and a decreased dispersion and increased density function height now describe the Cauchy distribution.

Because of the increased correlation between variates, the second classification problem defined by $\rho=0.98$ has a lower probability of error, a result which one would expect from observation of the shape of the decision boundary. Using the same techniques as just iterated, alpha is computed to be 0.347, and beta to be 0.130. In comparison, using techniques to be described in Chapter 4, the receiver for the original Gaussian distribution has alpha equal to 0.216 and beta equal to 0.053.

A significant increase in separation of the class one and class two Cauchy probability density functions is achieved by defining class one to have a correlation coefficient of equal magnitude but opposite sign to that of class two correlation coefficient. Very small values of

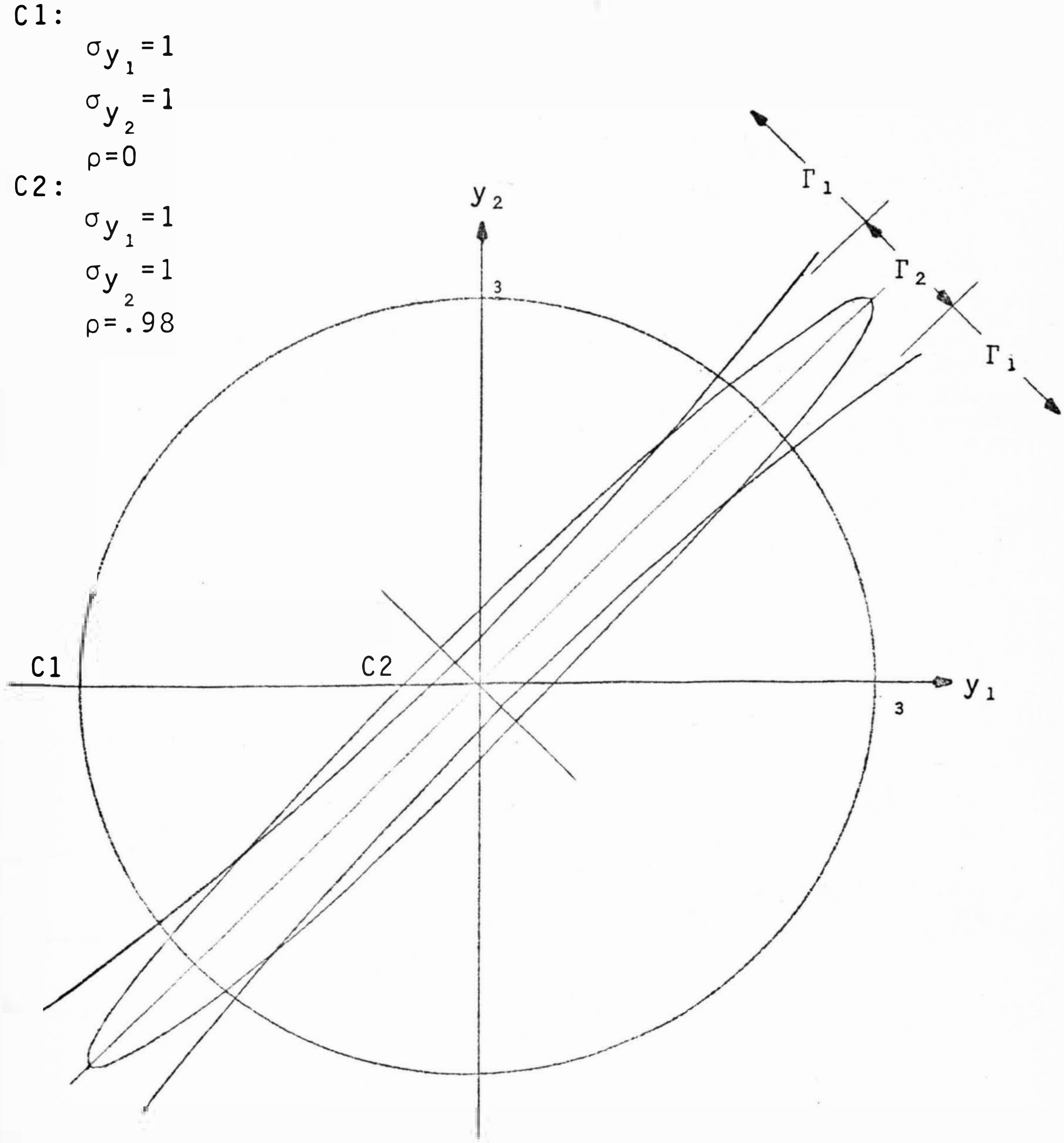


Figure 3.2. Decision boundary for Bayes' classification of two bivariate Gaussian distributions, ρ=0.98.

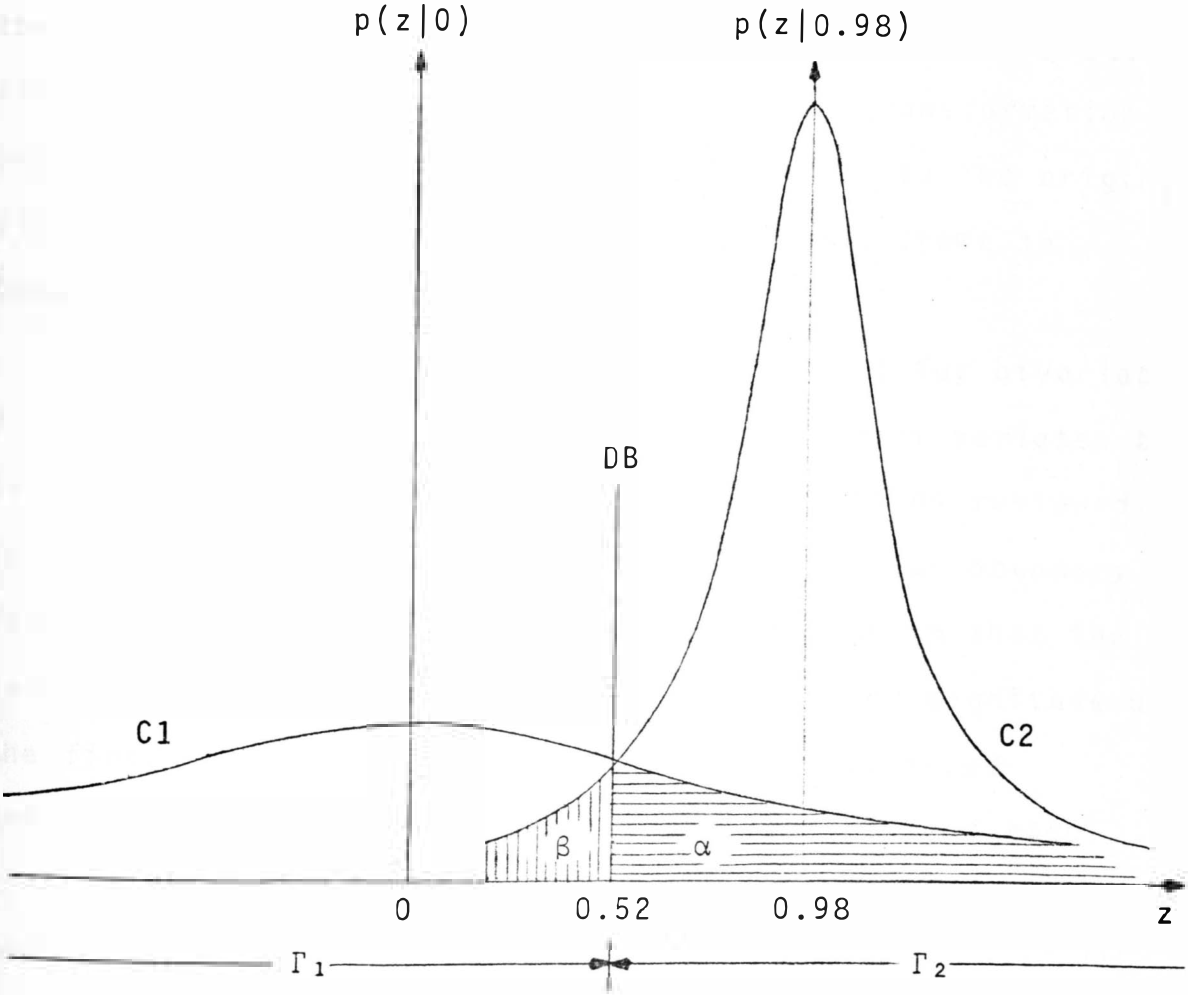


Figure 3.3. Decision boundary for Bayes' classification of two univariate Cauchy distributions; $\mu_1=0$, $\mu_2=0.98$.

alpha and beta are obtained, as can be seen in Figure 3.5. If one observes the class one and two bivariate Gaussian probability density functions illustrated in Figure 3.4, he notices as significant an improvement in the separation of the Gaussian density functions as was apparent in the Cauchy classification problem. Consequently, the transformation begins to appear as at best an approximation to the original distribution. More definite conclusions are drawn in Chapter 6.

Because the equations given in Table 2.1 for bivariate distributions consider only one class of random variates to be correlated, one must implement the techniques reviewed in section one of Chapter 2 to obtain a decision boundary for anti-correlated variates. The specification that the second correlation coefficient be equal to the magnitude of the first, but opposite in sign, plus the continued assumption of unity variance in both dimensions of each class permits considerable simplification of the decision boundary equation.

An interesting constraint is placed on the value of K for use in the resulting decision boundary, written below as:

$$\ln[K] = -3.92\, y_1 y_2 \quad (7)$$

If K equals unity, the decision boundary is simply the boundary formed by the $-y_1$, y_2 axes and the y_1, $-y_2$ axes. If the threshold K is, for example, 1.001, implying either slightly greater cost for an alpha error or slightly higher a priori probability of a class one data vector, (7) defines a hyperbola in the second and fourth quadrants and asymptotic to the $y_1 y_2$ axes. Values of K greater than unity cause the decision boundary to move outward from (0,0) as illustrated in Figure 3.4.

The problem of defining a decision boundary and evaluating alpha and beta for cases in which the correlation coefficients of class one and class two are not equal, and the variances of each class are unequal, is not considered in this thesis.

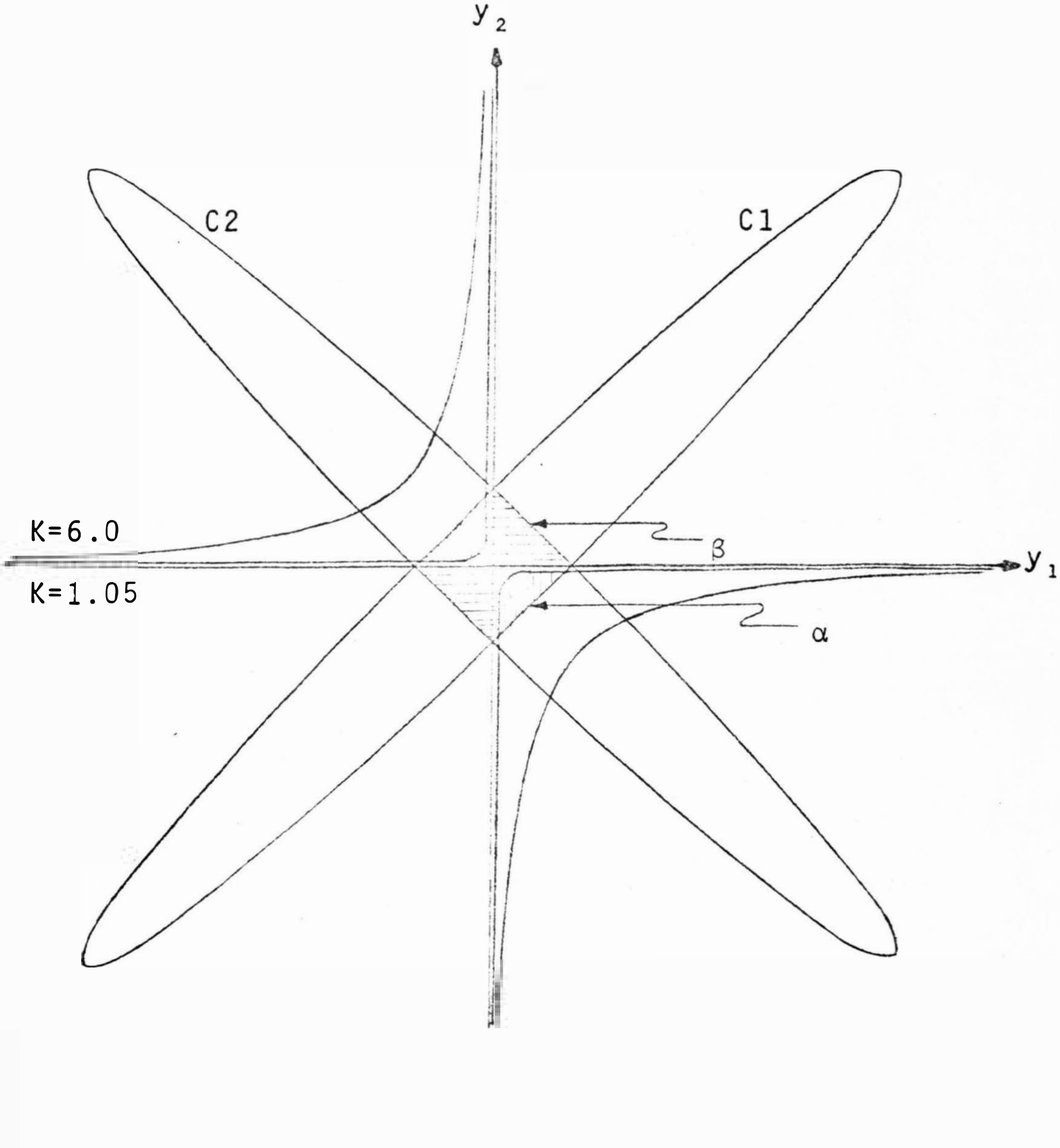


Figure 3.4. Two decision boundaries for two anticorrelated Gaussian distributions; ρ_1=0.98, ρ_2=-0.98.

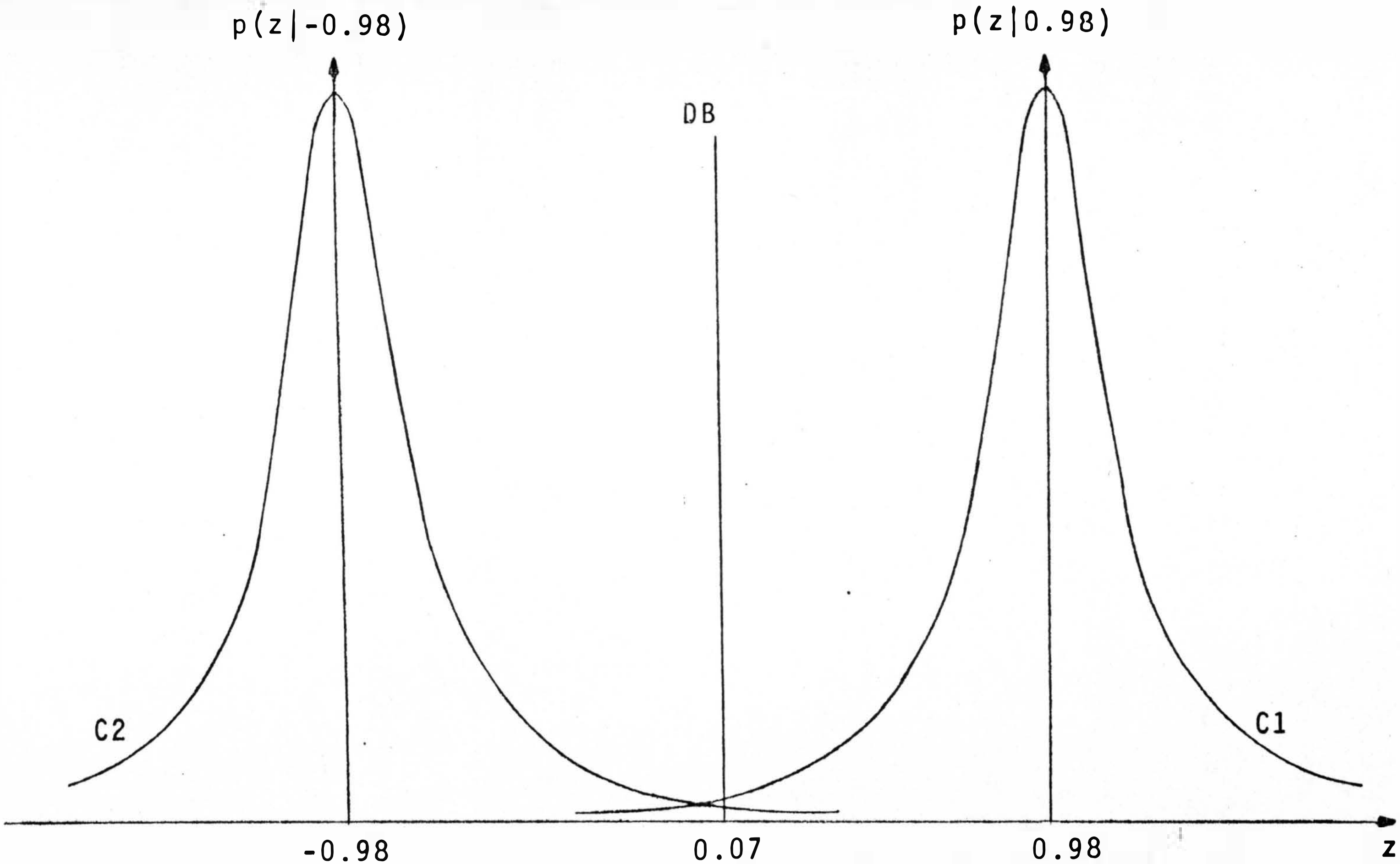


Figure 3.5. Decision boundary for two univariate Cauchy distributions; $\mu_1=.98$, $\mu_2=-.98$, $b_1=0.2$, $b_2=0.2$.

CHAPTER 4: COMPUTATION OF ALPHA AND BETA, RECEIVER OPTIMAL FOR BIVARIATE GAUSSIAN DISTRIBUTION

I. Rotational Transformation for Uncorrelation of Correlated Variates

Before detailing the calculation of classification errors alpha and beta, a method of transforming correlated random variates to uncorrelated variates by a rotation of coordinate axes is developed.

Later the requirement for both the decision boundary and the contours of constant probability density to be explicit functions of the variable will be shown. That is, both must be expressed in the form $y_1=f(y_2)$. As written in (1), the probability density function $p(y_1,y_2)$ contains a cross-product term y_1y_2 which must be eliminated. Comparing (2), which is the probability density function of two uncorrelated variates, with (1), one observes the only difference to be the magnitude of the multiplicative constants and the absence of the cross-product term. For clarity in notation of coordinate axes, the correlated random variates are denoted as y_1, y_2 and the uncorrelated variates are denoted as x_1, x_2.

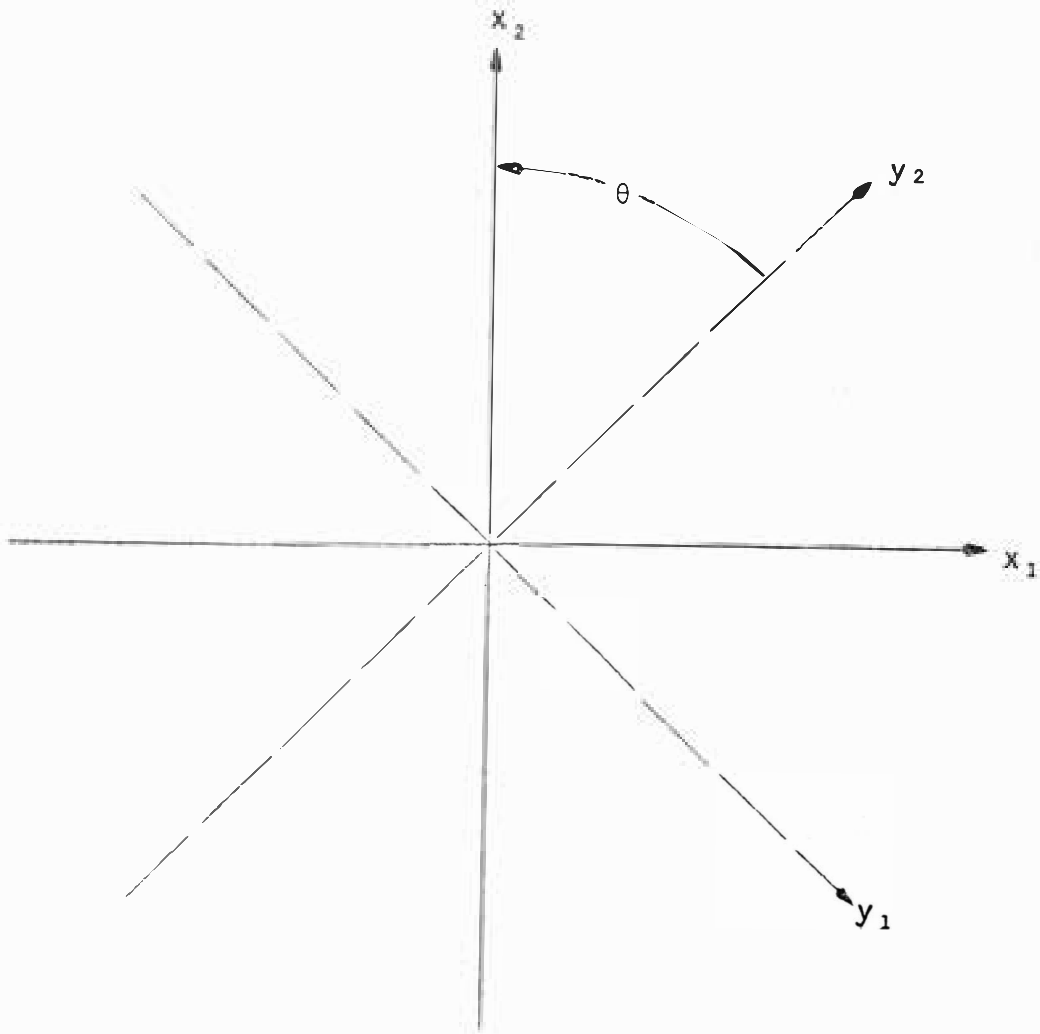
$$p(y_1,y_2)=\frac{1}{2\pi\sigma_{y_1}\sigma_{y_2}\sqrt{1-\rho^2}}\exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{y_1{}^2}{\sigma_{y_1}^2}-\frac{2\rho y_1 y_2}{\sigma_{y_1}\sigma_{y_2}}+\frac{y_2{}^2}{\sigma_{y_2}^2}\right)\right] \quad (1)$$

$$p(x_1,x_2)=\frac{1}{2\pi\sigma_{x_1}\sigma_{x_2}}\exp\left[-\frac{1}{2}\left(\frac{x_1{}^2}{\sigma_{x_1}^2}+\frac{x_2{}^2}{\sigma_{x_2}^2}\right)\right] \quad (2)$$

One must remember that for all discussions to this point, and continuing to Chapter 5, the mean of the probability density function is zero.

The contour of constant probability density is written as $y_1=f(y_2)$, or for uncorrelated variates $x_1=f(x_2)$, by transforming (1) into (2) using a rotation of coordinate axes. Davenport and Root [16] describe a transformation of uncorrelated variates to correlated; the inverse transformation is described here.

Figure 4.1 displays the pertinent information of the transformation. The angle θ is the radian measure of the distance through which the original axis y_2 must be rotated to coincide with the major axis, which is labeled the x_2 axis. The usual sign convention, of counter-clockwise increasing angles being positive, holds and must be strictly observed. For positive values of the correlation coefficient ρ, the y_2 axis lies in the second and fourth quadrants of the x_1x_2 coordinate system, and as a result the angle is negative. Equally important to remember, the declination



θ is defined from y_2 to x_2, positive in ccw direction.

$\theta=\phi$ for $\sigma_2>\sigma_1$

$\theta=-(\pi/2-\phi)$ for $\sigma_2<\sigma_1$

$$\phi=\frac{1}{2}\tan^{-1}\left(\frac{2\rho\sigma_{y_1}\sigma_{y_2}}{\sigma_{y_1}{}^2-\sigma_{y_2}{}^2}\right)$$

Figure 4.1. Definition of angle θ which describes a rotation transformation of two variates.

of the major axis of the ellipse formed by the contour of constant probability density is given by the angle θ, defined as

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\rho\sigma_{y_1}\sigma_{y_2}}{\sigma^2_{y_1}-\sigma^2_{y_2}}\right),$$

and related to the angle θ by

$$\theta = +\phi \qquad \text{for} \quad \sigma_{y_2} > \sigma_{y_1}$$

$$\theta = -(90°-\phi) \qquad \text{for} \quad \sigma_{y_2} < \sigma_{y_1} .$$

The random variates y_1 and y_2 are related to the random variates x_1 and x_2 by the rotational transformation

$$y_1 = x_1\cos\theta - x_2\sin\theta \tag{3}$$

$$y_2 = x_1\sin\theta + x_2\cos\theta . \tag{4}$$

With this background information, one is now able to write equations for the contour of constant probability density and the decision boundary which have no cross-product term. Although the cross-product is absent, one should not expect the correlation coefficient to be an unnecessary parameter, because the new multiplicative constants σ_{x_1}, σ_{x_2} are a function of ρ. Two different techniques for determining the new σ_{x_1} and σ_{x_2} are demonstrated in a later section of Chapter 4.

II. Transformation of the Contours of Constant Probability Density and of the Decision Boundary

The development begins with equations (10) and (11) of Chapter 2, which define the contour of constant probability density for any multiple of the standard deviation σ. Substituting the transformation equations given in the previous section for y_1 and y_2,

$$(1-\rho^2)\sigma^2 = \left(\frac{y_1{}^2}{\sigma_{y_1}^2} - \frac{2\rho y_1 y_2}{\sigma_{y_1}\sigma_{y_2}} + \frac{y_2{}^2}{\sigma_{y_2}^2}\right).$$

Collecting terms, the desired equation is written:

$$(1-\rho^2)\sigma^2 = \frac{1}{\sigma_{y_1}^2\sigma_{y_2}^2}\Big[(\cos^2\theta\sigma_{y_2}^2 - 2\rho\cos\theta\sin\theta\sigma_{y_1}\sigma_{y_2} + \sin^2\theta\sigma_{y_1}^2)x_1{}^2$$

$$+ (\sin^2\theta\sigma_{y_2}^2 + 2\rho\cos\theta\sin\theta\sigma_{y_1}\sigma_{y_2} + \cos^2\theta\sigma_{y_1}^2)x_2{}^2\Big]$$

$$\text{for } \sigma_{y_1} \neq \sigma_{y_2}, \qquad (5)$$

$$(1-\rho^2)\sigma^2 = \frac{1}{\sigma_{y_2}}\Big[(1-\rho)x_1{}^2 + (1+\rho)x_2{}^2\Big] \quad \text{for } \sigma_{y_1}=\sigma_{y_2}=\sigma_y. \qquad (6)$$

If a procedure is devised which permits the variances in the X domain to be calculated, the uncorrelated variate expression is used to compute the contours of constant probability density as

$$\sigma^2 = \left(\frac{x_1{}^2}{\sigma^2_{x_1}} + \frac{x_2{}^2}{\sigma^2_{x_2}} \right) . \tag{7}$$

A similar procedure is followed to determine an expression free of cross-product terms (that is, written in the X domain) for the decision boundary. The following equation is rewritten from Table 2.1.

$$2 \ln \left[\frac{\sigma_{2_{y_1}} \sigma_{2_{y_2}} (1-\rho^2)^{\frac{1}{2}} K}{\sigma_{1_{y_1}} \sigma_{1_{y_2}}} \right] = \left\{ y_1{}^2 \left(\frac{1}{\sigma^2_{1_{y_1}}} - \frac{1}{(1-\rho^2)\sigma^2_{2_{y_1}}} \right) \right.$$

$$\left. + \left(\frac{2\rho y_1 y_2}{(1-\rho^2)\sigma_{2_{y_1}} \sigma_{2_{y_2}}} \right) + y_2{}^2 \left(\frac{1}{\sigma^2_{1_{y_2}}} - \frac{1}{(1-\rho^2)\sigma^2_{2_{y_2}}} \right) \right\} \tag{8}$$

Substituting equations (3) and (4) for y_1 and y_2, respectively, and collecting terms one obtains the decision boundary

$$2 \ln \left[\frac{\sigma_{2y_1} \sigma_{2y_2} (1-\rho^2)^{\frac{1}{2}} K}{\sigma_{1y_1} \sigma_{1y_2}} \right] =$$

$$\left\{ (x_1^2 \cos^2\theta + x_2^2 \sin^2\theta) \left(\frac{1}{\sigma_{1y_1}^2} - \frac{1}{\sigma_{2y_1}^2 (1-\rho^2)} \right) \right.$$

$$+ (x_1^2 \sin^2\theta + x_2^2 \cos^2\theta) \left(\frac{1}{\sigma_{1y_2}^2} - \frac{1}{\sigma_{2y_2}^2 (1-\rho^2)} \right)$$

$$\left. + (x_1^2 - x_2^2) \cos\theta \sin\theta \left(\frac{2\rho}{\sigma_{1y_1} \sigma_{2y_2} (1-\rho^2)} \right) \right\}$$

$$\text{for } \sigma_{1y_1} \neq \sigma_{1y_2}, \quad \sigma_{2y_1} \neq \sigma_{2y_2}, \tag{9}$$

$$2 \ln \left[\frac{\sigma_2^2 (1-\rho^2)^{\frac{1}{2}} K}{\sigma_1^2} \right] = \left\{ (x_1^2 + x_2^2) \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2 (1-\rho^2)} \right) \right.$$

$$\left. + (x_1^2 - x_2^2) \left(\frac{\rho}{(1-\rho^2)\sigma_2^2} \right) \right\}$$

$$\text{for } \sigma_{1y_1} = \sigma_{1y_2} = \sigma_1, \quad \sigma_{2y_1} = \sigma_{2y_2} = \sigma_2. \tag{10}$$

In the preceding equations the extra digit in the variance subscript denotes the class to which it applies.

For example, in $\sigma^2_{1y_2}$, digit 1 denotes class one, y denotes the variance in the Y domain, and digit 2 denotes the axis to which the variance applies.

By constraining the correlation coefficient to be zero, and by defining equivalent variances in the X domain, a simplified expression is obtained for the decision boundary:

$$2 \ln \left[\frac{\sigma_{2_{X_1}} \sigma_{2_{X_2}} K}{\sigma_{1_{X_1}} \sigma_{1_{X_2}}} \right] = \left\{ X_1^2 \left(\frac{1}{\sigma^2_{1_{X_1}}} - \frac{1}{\sigma^2_{2_{X_1}}} \right) + X_2^2 \left(\frac{1}{\sigma^2_{1_{X_2}}} - \frac{1}{\sigma^2_{2_{X_2}}} \right) \right\}$$

$$\text{for} \quad \sigma_{1_{X_1}} \neq \sigma_{1_{X_2}} \text{ and } \sigma_{2_{X_1}} \neq \sigma_{2_{X_2}}, \tag{11}$$

$$2 \ln \left[\frac{\sigma_2{}^2 K}{\sigma_1{}^2} \right] = \left\{ (X_1{}^2 + X_2{}^2) \left(\frac{1}{\sigma_1{}^2} - \frac{1}{\sigma_2{}^2} \right) \right\}$$

$$\text{for } \sigma_{1_{X_1}} = \sigma_{1_{X_2}} = \sigma_1 \text{ and } \sigma_{2_{X_1}} = \sigma_{2_{X_2}} = \sigma_2. \tag{12}$$

The simplification between (9) and (11), and to a lesser degree between (10) and (12), is immediately obvious. Thus, a clear choice of two methods is presented, one more appealing than the other. However, the computation of the X domain variance adds computation to the second method, making the two choices approximately equal in terms of computational effort and time; but, one method may be easier to visualize than the other.

To emphasize an important fact, one must be aware that if uncorrelated Gaussian random variates are available initially then (7) should be used for determination of the contours of constant probability density and (11) or (12) used for computation of the decision boundary. If the variates are correlated, two methods of calculation are available. One, use equations (5) and (9) or (6) and (10) to obtain the decision boundary and the constant probability density contours; or, compute the X domain variances using the techniques presented in the next section, and use (7) and (11) or (12) to obtain the two desired equations.

III. Determination of the X Domain Variances

To use (7), (11), or (12) with a Gaussian distribution of correlated variates, new variances must be used which are functions of ρ, θ, σ_{y_1}, and σ_{y_2}. Two methods are presented, one based on statistical relations and the other derived by equating coefficients of the probability density functions for correlated and uncorrelated variates. The former will be discussed first.

Blachman [15] defines the parameters of the bivariate Gaussian distribution as

$$E[x_i] = \mu_i, \quad E[(x_i-\mu_i)^2] = \sigma_i^2 \qquad \text{and} \tag{13}$$

$$\rho = \frac{E[(x_1-\mu_1)\ (x_2-u_2)]}{\sigma_1\sigma_2} . \tag{14}$$

Because the means of both classes of Gaussian random variates are equal to zero, (13) and (14) reduce to

$$E[x_i^2] = \sigma_i^2 \qquad \text{and} \tag{15}$$

$$\rho = \frac{E[x_1 x_2]}{\sigma_1 \sigma_2} \tag{16}$$

The desired variances σ_{x_1} and σ_{x_2} are calculated by computing the expectation of the inverse of the transformation written in (3) and (4),

$$x_1 = y_1 \cos\theta + y_2 \sin\theta \tag{17}$$

$$x_2 = -y_1 \sin\theta + y_2 \cos\theta \tag{18}$$

substituting x_1, x_2 into (15),

$$\sigma_{x_1}^2 = E[(y_1 \cos\theta + y_2 \sin\theta)^2].$$

Expanding,

$$\sigma_{x_1}^2 = \cos^2\theta E[y_1{}^2] + 2\cos\theta\sin\theta E[y_1 y_2] + \sin^2\theta E[y_2{}^2];$$

replacing $E[y_i{}^2]$ by (15), and $E[x_1 x_2]$ by (16) the desired result is obtained:

$$\sigma_{x_1}^2 = \cos^2\theta\sigma_{y_1}^2 + 2\cos\theta\sin\theta\rho\sigma_{y_1}\sigma_{y_2} + \sin^2\theta\sigma_{y_2}^2 . \tag{19}$$

Similarly,

$$\sigma_{x_2}^2 = \sin^2\theta\sigma_{y_1}^2 - 2\cos\theta\sin\theta\rho\sigma_{y_1}\sigma_{y_2} + \cos^2\theta\sigma_{y_2}^2 \ . \tag{20}$$

The dependence of σ_x on ρ and σ_y, apparently lacking in (7), (11), and (12) is, in fact, present; the relationship of ρ, θ and σ_y with σ_x, not obvious previously, is explicitly stated in (19) and (20).

As mentioned previously, a second method exists for deriving the X domain variances, that of equation of coefficients. Refer to (5); the expression on the right is the major portion of the exponent term of the probability density function. Written in its entirety, the density function includes two other constants in addition to the exponent.

$$p(y_1,y_2) = \frac{1}{2\pi\sigma_{y_1}\sigma_{y_2}\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)\sigma_{y_1}^2\sigma_{y_2}^2}\left(ax_1{}^2 + bx_2{}^2\right)\right] \tag{21}$$

where $a=\cos^2\theta\sigma_{y_2}^2 - 2\rho\cos\theta\sin\theta\sigma_{y_1}\sigma_{y_2} + \sin^2\theta\sigma_{y_1}^2$ and

$b=\sin^2\theta\sigma_{y_2}^2 + 2\rho\cos\theta\sin\theta\sigma_{y_1}\sigma_{y_2} + \cos^2\theta\sigma_{y_1}^2$.

The density function for uncorrelated variates, written before as (2) but repeated below is

$$p(x_1,x_2) = \frac{1}{2\pi\sigma_{x_1}\sigma_{x_2}} \exp\left[-\frac{1}{2}\left(\frac{x_1{}^2}{\sigma_{x_1}^2} + \frac{x_2{}^2}{\sigma_{x_2}^2}\right)\right] \ . \tag{22}$$

To obtain the new variance σ_X, simply equate the coefficient of $x_1{}^2$ and $x_2{}^2$ of the two density functions.

$$\sigma_{x_1}^2 = \left(\frac{\sigma_{y_1}^2 \sigma_{y_2}^2 \;(1-\rho^2)}{\cos^2\theta\sigma_{y_2}^2 - 2\rho\cos\theta\sin\theta\sigma_{y_1}\sigma_{y_2} + \sin^2\theta\sigma_{y_1}^2}\right) \tag{23}$$

$$\sigma_{x_2}^2 = \left(\frac{\sigma_{y_1}^2 \sigma_{y_2}^2 \;(1-\rho^2)}{\sin^2\theta\sigma_{y_2}^2 + 2\rho\cos\theta\sin\theta\sigma_{y_1}\sigma_{y_2} + \cos^2\theta\sigma_{y_1}^2}\right) \tag{24}$$

Although not obvious by inspection, (19) is equal to (23) and (20) is equal to (24).

For the case considered in this thesis, $\sigma y_1 = \sigma y_2 = 1.0$, either set of equations will yield the result

$$\sigma_{x_1}^2 = (1+\rho), \qquad \sigma_{x_2}^2 = (1-\rho) \; . \tag{25}$$

With the information presented, one not only has two methods of computing the contours of constant probability density and decision boundary, but also has, should the second method be chosen, two equations for the computation of the variances σ_{x_1} and σ_{x_2}.

Figure 4.2 illustrates the transformation of correlated variates to uncorrelated variates for an arbitrarily chosen probability density function.

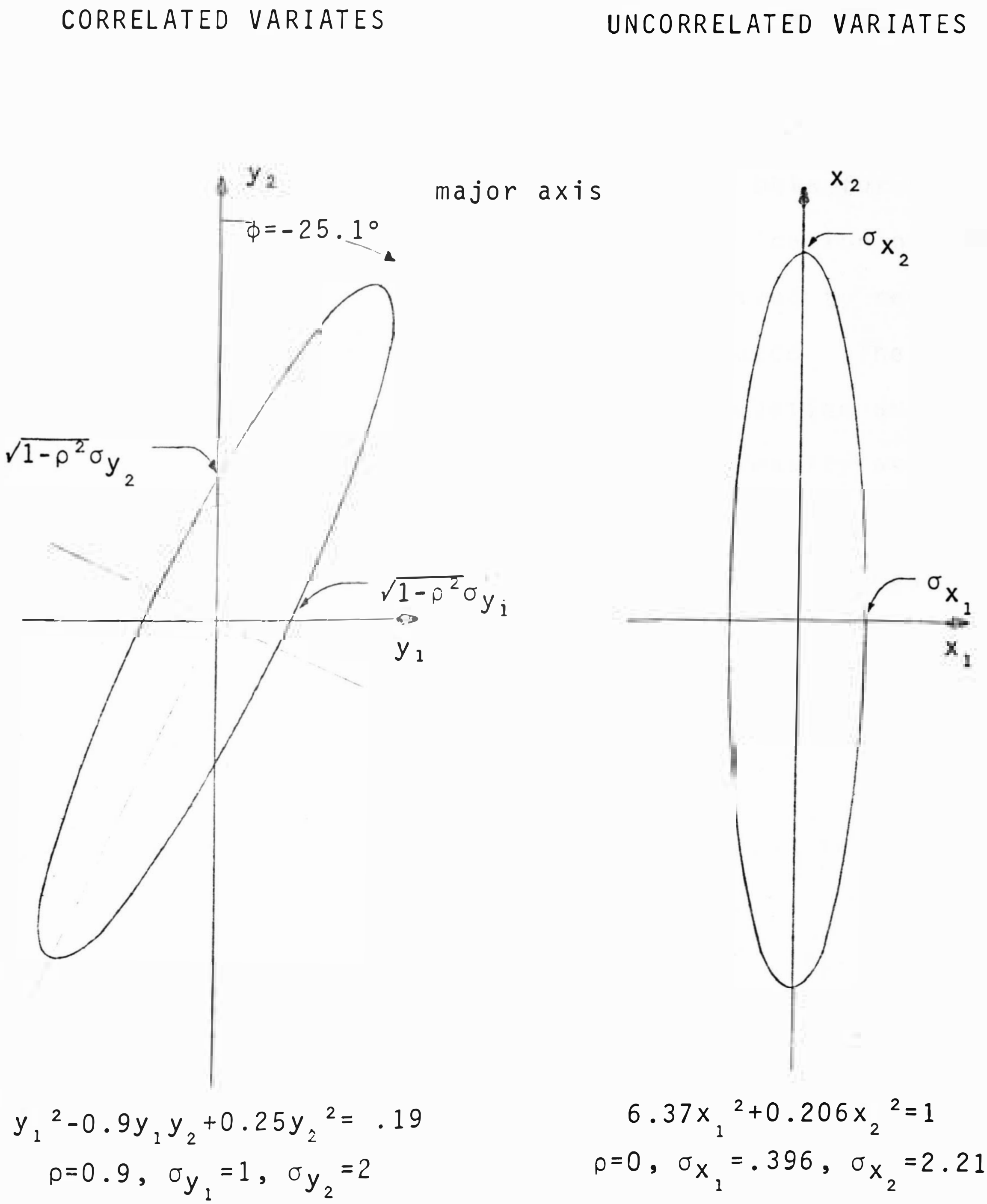


Figure 4.2. One-sigma contour of Gaussian probability density function for correlated and uncorrelated variates.

IV. Computation of Alpha and Beta

Now that expressions have been obtained for the decision boundary and contours of constant probability density in which x_1 and x_2 are separated, it is possible to proceed with the computation of alpha and beta for the original two-class bivariate Gaussian classification problem. The advantage which symmetry affords is utilized to reduce the range over which the integrals are evaluated. The abbreviations DB for the decision boundary equation and PDF for the contour of constant probability density are used to provide a clearer notation.

Substitute into the integral expression for alpha, which is

$$\alpha = \int_{\Gamma_2} p(\underline{X}|\underline{\theta}_1)\,d\underline{X},$$

the Gaussian probability density function to obtain

$$\alpha = \frac{4}{2\pi\sigma_{1x_1}\sigma_{1x_2}} \int_0^{\infty}\int_0^{DB} \exp\left[\frac{-x_1^{\,2}}{2\sigma_{1x_1}^2} - \frac{x_2^{\,2}}{2\sigma_{1x_2}^2}\right] dx_1 dx_2 .$$

Let $t^2 = x_1^{\,2}/2\sigma_{1x_1}^2$; when $x_1 = DB$, $t = DB/\sqrt{2}\,\sigma_{1x_1}$. Therefore,

$$\alpha = \left(\frac{\sqrt{2}}{\sqrt{\pi}}\,\frac{1}{\sigma_{1x_2}}\right)\int_0^{\infty}\int_0^{DB/\sqrt{2}\sigma_{1x_1}} \left\{\frac{2}{\sqrt{\pi}} \exp\,(-t^2)\right\} \exp\left[\frac{-x_2^{\,2}}{2\sigma_{1x_2}^2}\right] dt\, dx_2 .$$

Using the equation

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-x^2)\, dx,$$

and integrating only to the three sigma limit on the outer integral to reduce computation time, the following integral is obtained:

$$\alpha = \left(\frac{2}{\pi}\frac{1}{\sigma_{1_{X_2}}}\right) \int_0^{3\sigma_{1_{X_2}}} \mathrm{erf}(DB/\sqrt{2}\,\sigma_{1_{X_1}}) \exp\left[\frac{-X_2{}^2}{2\sigma^2_{1_{X_2}}}\right] dx_2 .$$

In the expressions for alpha and beta the X domain variances are written with subscripts to designate the class and the axis to which the variance applies, $\sigma^2_{1_{X_2}}$ being the variance along the x_2 axis of the class one probability density function. For the specific problem discussed in this thesis, $\sigma_{1_{X_1}} = \sigma_{1_{X_2}} = 1.00$ and the expression for alpha reduces to

$$\alpha = \frac{2}{\pi} \int_0^{3\sigma_X} \mathrm{erf}(DB/\sqrt{2}) \exp\left[\frac{-X_2{}^2}{2}\right] dx_2 .$$

Integrate by using the trapezoidal rule or Simpson's rule. The upper limit of three sigma is valid because a negligible volume is contributed outside of the boundary. Either (9) or (11) may be used for the expression abbreviated DB in the argument of the error function.

Beta must be written as the sum of two integrals, or as the integral of the difference between two error functions. For ease in programming the numerical integration, the former is used. In the following equation the probability density function is expressed in terms of uncorrelated variates, indicating that the rotational transformation has already been performed and that the X domain variance is known or can be calculated easily.

$$\beta = \int_{\Gamma_1} p(\underline{X}|\underline{\theta}_2)\, d\underline{X}$$

Substitute the Gaussian probability density function into the integral equation for beta to obtain

$$\beta = 1.0 - \frac{4}{2\pi\sigma_{X_1}\sigma_{X_2}}\left\{\int_0^{2.6}\int_0^{PDF} \exp\left[\frac{-x_1^2}{2\sigma_{X_1}^2} - \frac{x_2^2}{2\sigma_{X_2}^2}\right]dx_1 dx_2\right.$$

$$\left. + \int_{2.6}^{4.14}\int_0^{PDF} \exp\left[\frac{-x_1^2}{2\sigma_{X_1}^2} - \frac{x_2^2}{2\sigma_{X_2}^2}\right]dx_1 dx_2\right\}$$

Let $t^2 = x_1^2/2\sigma^2_{2X_1}$; when $x_1 = PDF$, $t = PDF/\sqrt{2}\sigma_{2X_1}$.

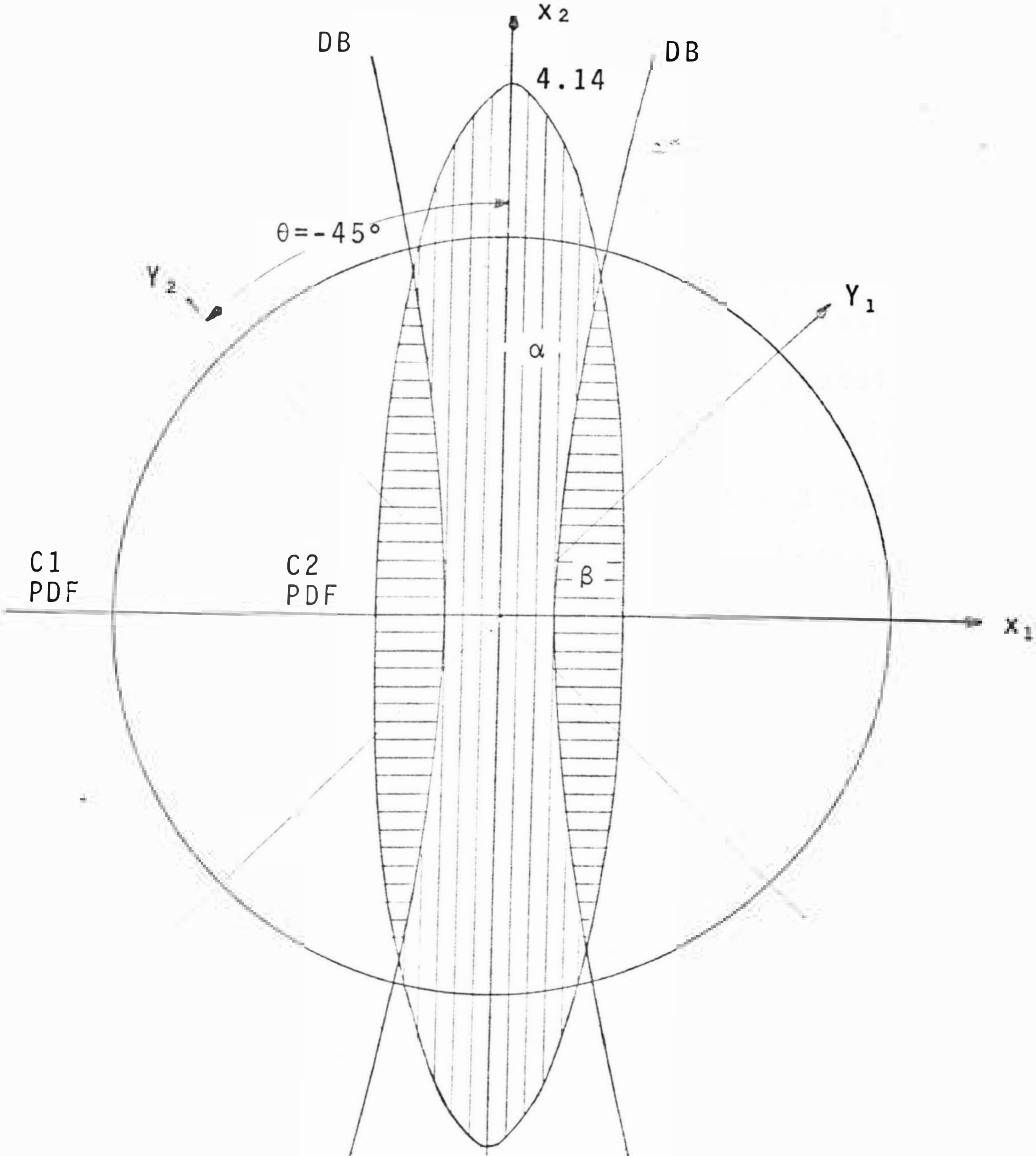
After substituting the change of variables and then recognizing the inner integral defines the error function, the desired result is obtained:

$$\beta = 1.0 - \frac{2}{\pi} \frac{1}{\sigma_{2_{X_2}}} \left\{ \int_{0}^{2.6} \operatorname{erf}(PDF/\sqrt{2}\sigma_{2_{X_1}}) \exp\left[\frac{-X_2^{\,2}}{2\sigma^2_{2_{X_2}}}\right] dX_2 \right.$$

$$\left. + \int_{2.6}^{4.14} \operatorname{erf}(PDF/\sqrt{2}\sigma_{2_{X_1}}) \exp\left[\frac{-X_2^{\,2}}{2\sigma^2_{2_{X_2}}}\right] dX_2 \right\} .$$

Variances $\sigma^2_{2_{X_1}}$ and $\sigma^2_{2_{X_2}}$ are computed from (19) and (20), and the equation for the PDF limit is (7). One must numerically integrate the above expression using the trapezoidal rule or Simpson's rule to obtain the value for beta. The upper limit on the two outer integrals must be obtained from knowledge of the decision boundary and the constant probability density contours. Upper limit 2.6 is obtained from the intersection of the two equations (7) and (9), and 4.14 is the three sigma limit for the class two probability density function. Refer to Figure 4.3 for a pictorial representation of the two Gaussian probability density functions.

The statement has been made that either the trapezoidal rule or Simpson's rule may be used for the numerical integration. Initially, the trapezoidal rule was used with an incremental change of 0.01 in the variable of integration. To check the answer, the integration was accomplished using Simpson's rule with an incremental change of 0.005 in the variable of integration. Essentially the same values



Prior to transformation:

$\sigma_{y_1}^2=1$ $\sigma_{y_1}=1$

$\sigma_{y_2}^2=1$ $\sigma_{y_2}=1$

$\rho=.90$

Following transformation:

$\sigma_{x_1}^2=0.1$ $\sigma_{x_1}=.316$

$\sigma_{x_2}^2=1.9$ $\sigma_{x_2}=1.38$

$\rho=0$

Figure 4.3. Two-class bivariate Gaussian distribution after rotational transformation of variates.

for alpha and beta were obtained. For the Gaussian distribution having correlation coefficient ρ equal to 0.90, alpha is 0.370 and beta is 0.122. These values are lower than the corresponding classification errors of the Cauchy distribution: alpha is 0.394 and beta is 0.216. Comparison of alpha and beta for the two distributions shows more errors are made if the transformed distribution is classified than if the original bivariate Gaussian distribution is classified.

A review of results discussed in Chapter 3 indicates that the original distribution, bivariate Gaussian with ρ=0.98, is again classified with fewer errors than the transformed distribution, which is univariate Cauchy. Alpha for the original distribution is 0.216, compared with 0.347 for the transformed distribution; and beta for the original distribution is 0.053, compared with 0.130 for the transformed distribution. A tabular listing of the above results and comments on their implied meaning is included in Chapter 6.

To provide assurance that the numerical integration was accomplished correctly, a check, which is described in the next section, was performed.

V. Volume Under the Probability Density Function

Because the volume under the probability density function must equal unity, the author decided to integrate a standard Gaussian density function within the boundary formed by the three sigma limits, and to compare that answer to unity. An answer of approximately 0.99 was expected, because a small amount of volume is excluded if the three sigma limits are used. Using Simpson's rule for the numerical integration, an answer of 0.989 was obtained.

Next, attention was focused on the volume within the boundaries formed by the one and two sigma limits. One may recall that the area under the probability density function for a random variable normally distributed is 0.6826, 0.9544, and 0.9974 for the one, two, and three sigma limits respectively. For a two-dimensional (bivariate) case, one must speak in terms of volume rather than area.

The volume computed within the one sigma limit, for instance, totalled 0.393. Other computed volumes are shown in Figure 4.4. The numbers to the right of the figure denote the volume within the one, two, and three sigma limits. They are cumulative values, the volume inside the two sigma boundary naturally including that within the one sigma boundary. Numbers to the left of the figure represent the volume within the truncated ellipse formed by the three sigma limit along x_1, and the one, two, or three sigma limit along x_2. Had the limit of the integration in the x_1 direction been the fifth multiple of sigma instead of the third, the numbers on the left would equal those already given for the area under a one-dimensional probability density function.

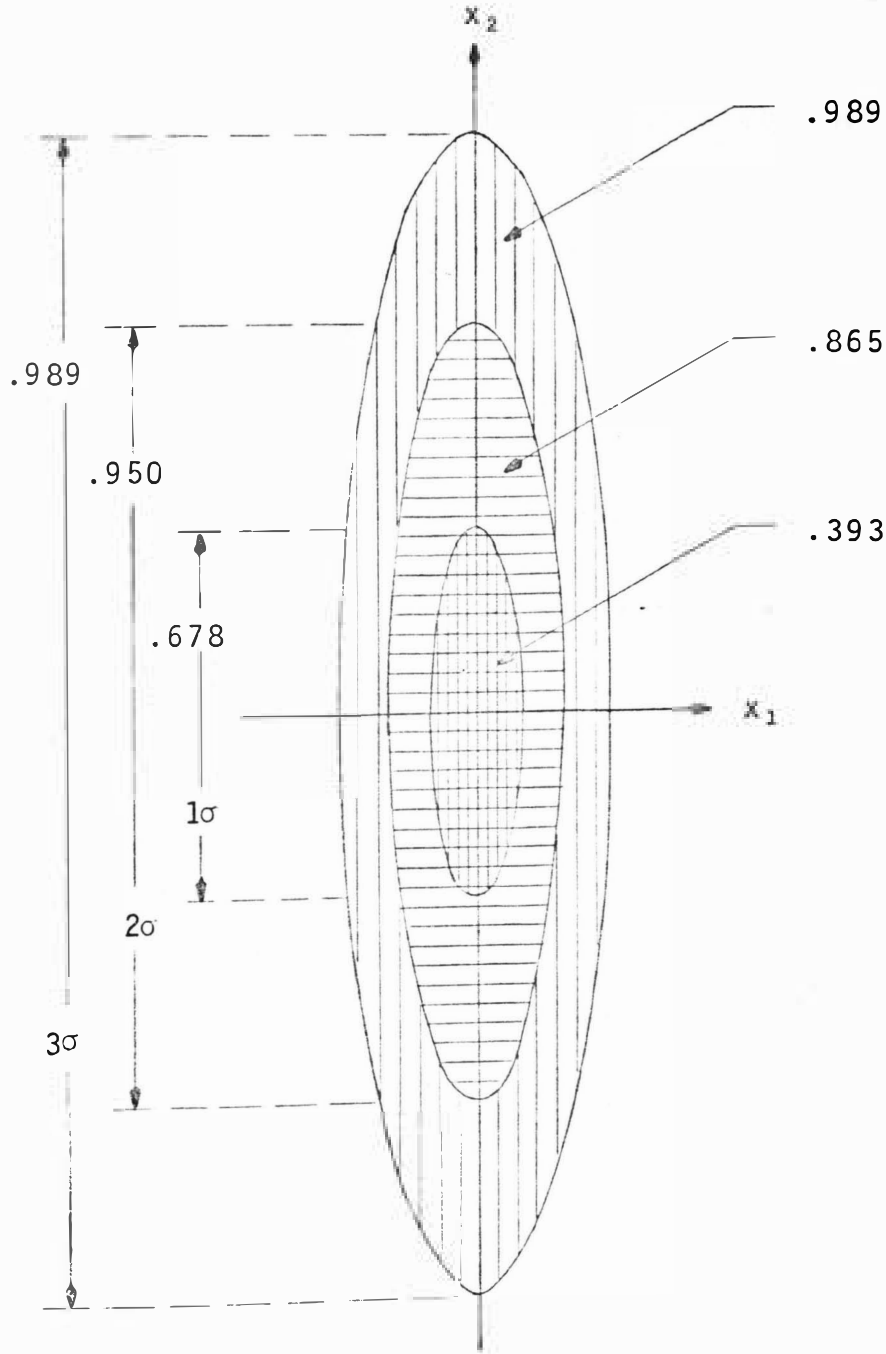


Figure 4.4. Volume contained under the Gaussian probability density function, within the 1, 2, and 3 σ limits.

The following proof demonstrates that $(0.6826)^2$ is an approximation to the true volume. The approximate volume is obtained analytically by evaluating the volume integral not within the circular boundary formed by the one sigma contour of constant probability density, but within a unit square. A standardized bivariate Gaussian distribution is assumed. Both the exact and approximate limits are indicated in Figure 4.5.

$$V = \frac{4}{2\pi} \int_0^1 \int_0^1 \exp \frac{-x_1{}^2}{2} \frac{-x_2{}^2}{2} \, dx_1 dx_2$$

Let $u^2 = x_1{}^2/2$ and $v^2 = x_2{}^2/2$.

$$V = \frac{2}{\pi} \int_0^{1/\sqrt{2}} \frac{2}{\pi} \int_0^{1/\sqrt{2}} \exp(-u^2) \exp(-v^2) \, du dv$$

Both integrals define error functions of argument $(1/\sqrt{2})$,

$$V = \mathrm{erf}(1/\sqrt{2}) \, \mathrm{erf}(1/\sqrt{2}) = (0.6826)^2 = 0.466$$

Because the solution includes additional volume, 0.466 is an approximation to the true volume.

After the preceding approximation had been derived and the volumes given in Figure 4.4 computed, a reference was found supporting the results reported here. Martin [17] calculated the volume within the one, two, and three sigma boundary for a bivariate Gaussian distribution having equal

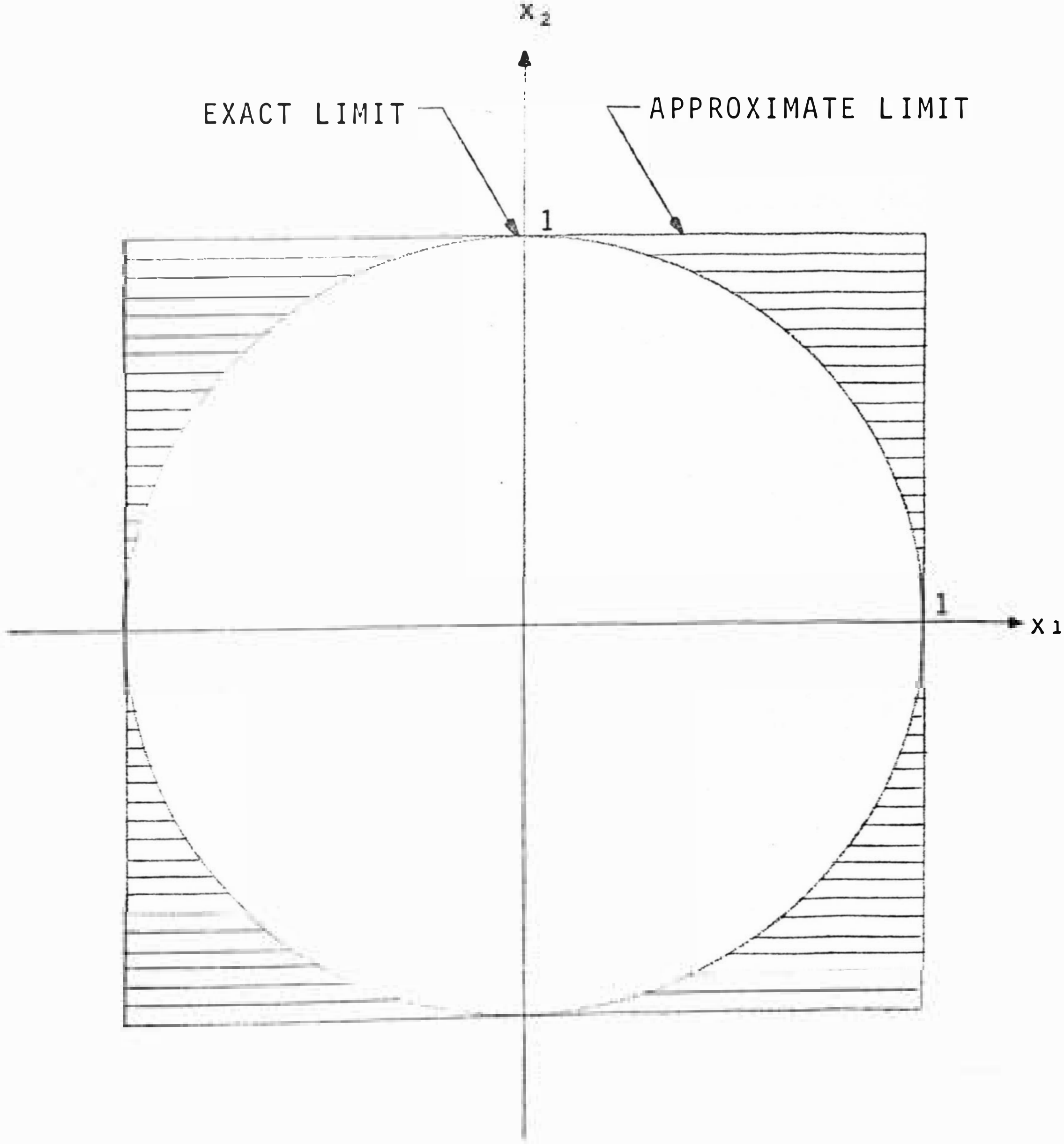


Figure 4.5. Exact and approximate boundary for volume integration.

variances. His results are also valid for distributions having unequal variances, such as the example in Figure 4.4. Martin reported volumes under the probability density function, within the boundaries specified, of 0.393, 0.865, and 0.989. The volumes listed to the right of the figure in Figure 4.4 are exactly the same as his results.

CHAPTER 5: SUMMARY OF OTHER INVESTIGATIVE STUDIES

I. Classification of Gaussian and Cauchy Variates by a Bayes' Receiver Optimal for a Gaussian Distribution

A discussion of functions of two random variates, by Papoulis [4], reveals that the probability density function of the random variate formed by the quotient of two jointly Gaussian random variates y_1 and y_2 is Cauchy distributed. This same result is the source of problem 4.8 in Hogg and Craig [18]. The variances of the bivariate Gaussian distribution determine the mean. Correlated Gaussian variates of mean zero and variances σ_1, σ_2 transform to Cauchy random variate of mean $\rho\sigma_1/\sigma_2$ and shape $b=\sqrt{1-\rho^2}\ \sigma_1/\sigma_2$.

As illustrated in previous figures, the Cauchy distribution is similar in shape to the Gaussian distribution, having less area under the curve at a distance of one sigma from the mean, but having greater area under the curve at a distance from the mean of three sigma and further. This similiarity prompted a study to determine how well a receiver defined for Gaussian random variates would classify variates having a distribution of the same mean and equivalent peak height as the Gaussian distribution.

Figure 5.1 shows the closeness of fit between the two distributions for one value of σ and b; other graphs have been drawn which indicate that the approximation varies little for other values of σ. An alternative approximation

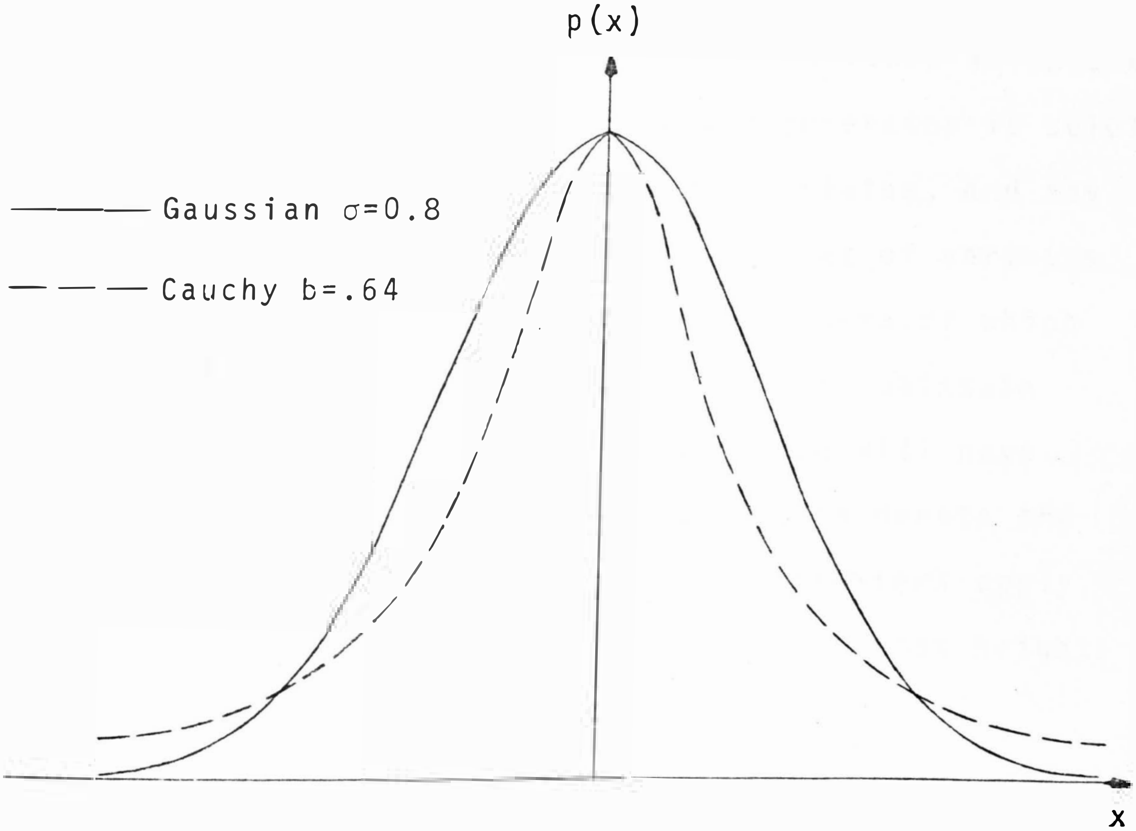


Figure 5.1. Illustration of similarity between Gaussian and Cauchy probability density functions.

allowing the Cauchy density function to project higher than the Gaussian at the mean of the probability density function, to permit somewhat balancing errors of approximation, was considered and rejected.

Simulation was performed with bivariate distributions to add some generality to the study. In the following description of the parameters and equations required for the simulation, keep in mind that one set of variates which are obtained from the Gaussian random number generator is solely for the generation of two Cauchy random variates, and may likely have a different variance than the set of variates obtained from the Gaussian random number generator which is to be classified by the receiver. To help maintain the concept of this difference, the variance will have subscripts c and g in addition to 1 and 2, to denote the class and the distribution to which the parameters apply.

First, for class one variates equate the peak heights of the two types of distributions:

$$\frac{1}{2\pi\sigma_{g_1}\sigma_{g_2}} = \left(\frac{1}{b\pi}\right)^2 \text{, where } b = \sigma_{c_1}/\sigma_{c_2} . \tag{1}$$

Assume that $\sigma_{g_1} = \sigma_{g_2}$. Therefore,

$$\frac{1}{2\pi\sigma_g^2} = \left(\frac{\sigma_{c_2}^2}{\pi^2\sigma_{c_1}^2}\right). \tag{2}$$

In the preceding equation σ_g is specified, σ_{c_1} is selected, and σ_{c_2} is calculated.

To obtain the one-dimensional Cauchy distribution which has a maximum height specified by $1/2\pi\sigma_{g_1}\sigma_{g_2}$ it is necessary to generate bivariate Gaussian random variates specified by zero mean and variance b as determined from (1) and (2). A bivariate Cauchy distribution of independent variates centered at (0,0) is obtained by computing two uncorrelated bivariate Gaussian random variates.

Class two variates, having a mean of (μ,μ) are more difficult to formulate but not to generate. The equations required are written below and in (4).

$$\frac{1}{2\pi\sigma_{g_1}\sigma_{g_2}} = \left(\frac{1}{b\pi}\right)^2 \quad \text{where } b = \frac{\sqrt{1-\rho^2}\,\sigma_{c_1}}{\sigma_{c_2}}$$

Assume $\sigma_{g_1} = \sigma_{g_2}$. (Variance $\sigma^2_{g_1}$ and $\sigma^2_{g_2}$ may be different from the previous values of $\sigma^2_{g_1}$ and $\sigma^2_{g_2}$). Substitution yields

$$\frac{1}{2\pi\sigma^2_g} = \left(\frac{\sigma^2_{c_2}}{\pi^2\sigma^2_{c_1}(1-\rho^2)}\right). \tag{3}$$

$$\rho\frac{\sigma_{c_1}}{\sigma_{c_2}} = \mu. \tag{4}$$

Substituting $\rho=\mu\sigma_{c_2}/\sigma_{c_1}$ into (3), and calling $1/2\pi\sigma_g^2=h$, one obtains

$$\sigma_{c_2} = \left(\frac{h\pi^2}{1+h\pi^2\mu^2}\right)^{\frac{1}{2}}\sigma_{c_1} \qquad = C\ \sigma_{c_1}\ . \tag{5}$$

To satisfy (4), the following equality must hold:

$$\rho = \frac{\mu C\sigma_{c_1}}{\sigma_{c_1}} = \mu C\ . \tag{6}$$

To obtain the required distribution parameters, compute h and select μ equal to the mean of the given Gaussian distribution; select σ_{c_1}, then compute σ_{c_2} and ρ. A bivariate Cauchy distribution centered at (μ,μ) is obtained by generating two correlated bivariate Gaussian distributions defined by σ_{c_1}, σ_{c_2}, and ρ.

The decision boundary which classifies the Cauchy or Gaussian random variates is written in Table 2.1; the receiver assumes a priori independent Gaussian variates when, in fact, it is called upon to classify independent Cauchy variates as well as independent Gaussian variates. For convenience the decision boundary is repeated below:

$$2\sigma_1^2\sigma_2^2 \ln\left[\frac{K\sigma_2^2}{\sigma_1^2}\right] + 2\mu^2\sigma_1^2 = (\sigma_2^2-\sigma_1^2){x_1}^2 + (\sigma_2^2-\sigma_1^2){x_2}^2 + 2\mu\sigma_1^2(x_1+x_2), \tag{7}$$

where σ_1^2 denotes the variance of class one, and σ_2^2 denotes the variance of class two. The variance $\sigma_{1_{x_1}}^2=\sigma_{1_{x_2}}^2$ is equal to σ_1^2; and $\sigma_{2_{x_1}}^2=\sigma_{2_{x_2}}^2$ is equal to σ_2^2.

If the variances of the two classes are equal as well, (7) reduces to equation 3-35 in Hancock and Wintz [19], which is the correct result:

$$x_1+x_2 = \mu + \frac{\sigma^2 \ln K}{\mu} .$$

Naylor [20] describes a technique for generating multivariate Gaussian random variates. The general equation is written below, but only the final result of the necessary matrix algebra is included. The reader is referred to Naylor for the complete discussion, including the set of recursive formulas which permit computation of the elements of the $\underline{C}$ matrix.

In general,

$$\underline{X} = \underline{C}\ \underline{z} + \underline{\mu} .$$

For generation of the bivariate Gaussian distribution the equation

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 z_1 \\ \sigma_2 z_2 \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

is used, and for generation of the Cauchy distribution the equation

$$U = \frac{x_1}{x_2}, \quad \text{where} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 z_1 \\ \rho\sigma_2 z_1 + \sigma_2\sqrt{1-\rho^2}\ z_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

is used. The variables z_1 and z_2 are standard normal variates generated by a Fortran subroutine.

Three different problems are considered, one in which the two distributions are widely separated, one in which some overlap between classes exists, and a third in which one distribution completely overlaps the other. Figure 5.2 illustrates the classification problem in a block diagram. In Figure 5.3 the amount of overlap of the two classes of data is shown for the three classification problems. Table 5.1 lists the results of the computer simulation; the confusion matrix lists the number of data vectors correctly and incorrectly classified. The top and side headings list the two classes; the number of data vectors assigned to each class is read at the intersection of the two headings. For example, in problem one class one Gaussian distributed data vectors were classified as class one 125 times, and classified as class two 0 times.

II. Transformation of Random Variates

Wozencraft and Jacobs [21] detail, in Chapter 2, transformations of one variable by addition of a constant, multiplication by a constant, and by squaring the variate. A selected transformation of two variates is treated by making a second change of variables, to polar coordinates. In the appendix of Chapter 2, Wozencraft and Jacobs [22] illustrate a reversible transformation of multivariate functions. However, they fail to strongly emphasize the

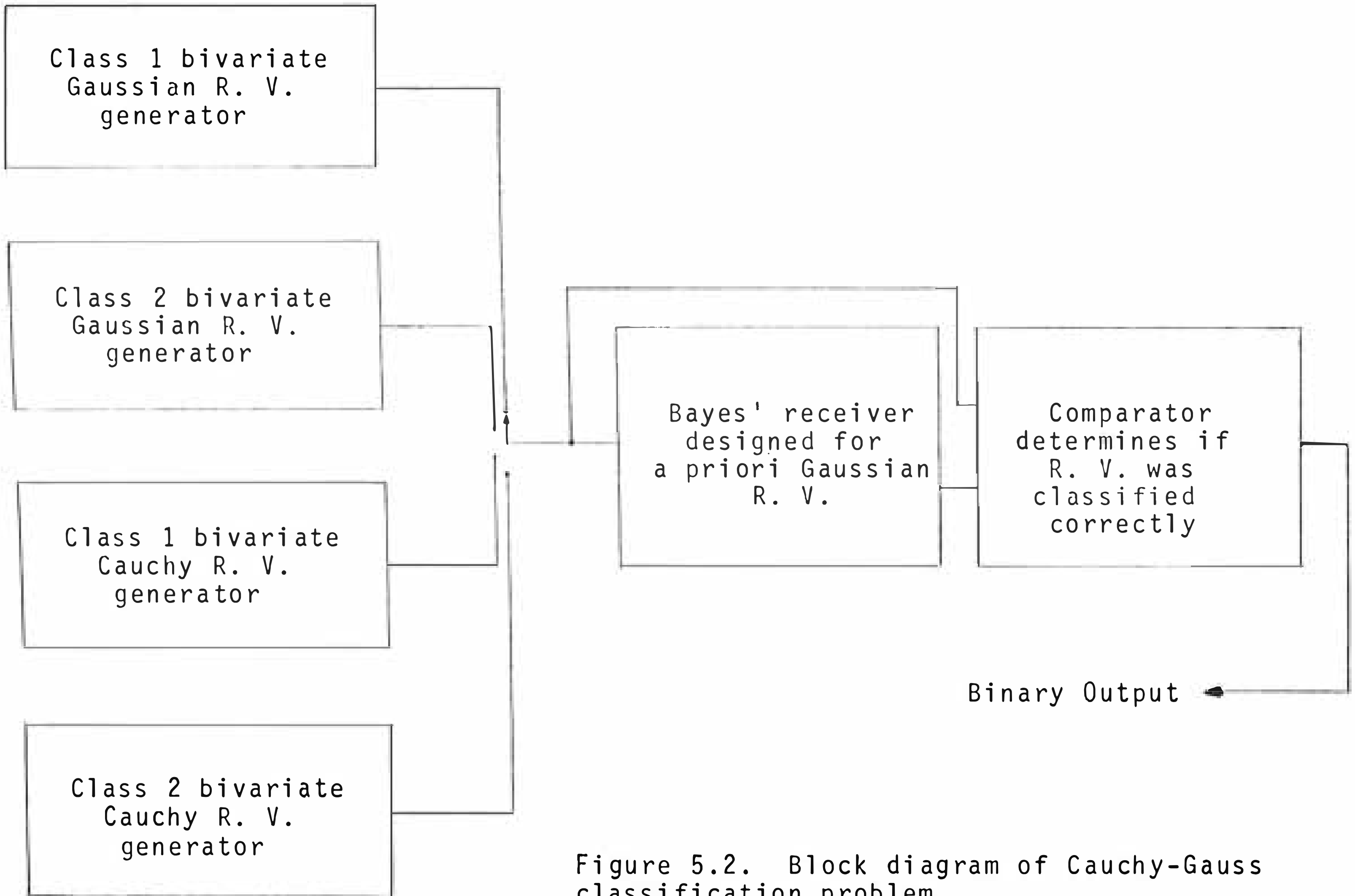


Figure 5.2. Block diagram of Cauchy-Gauss classification problem.

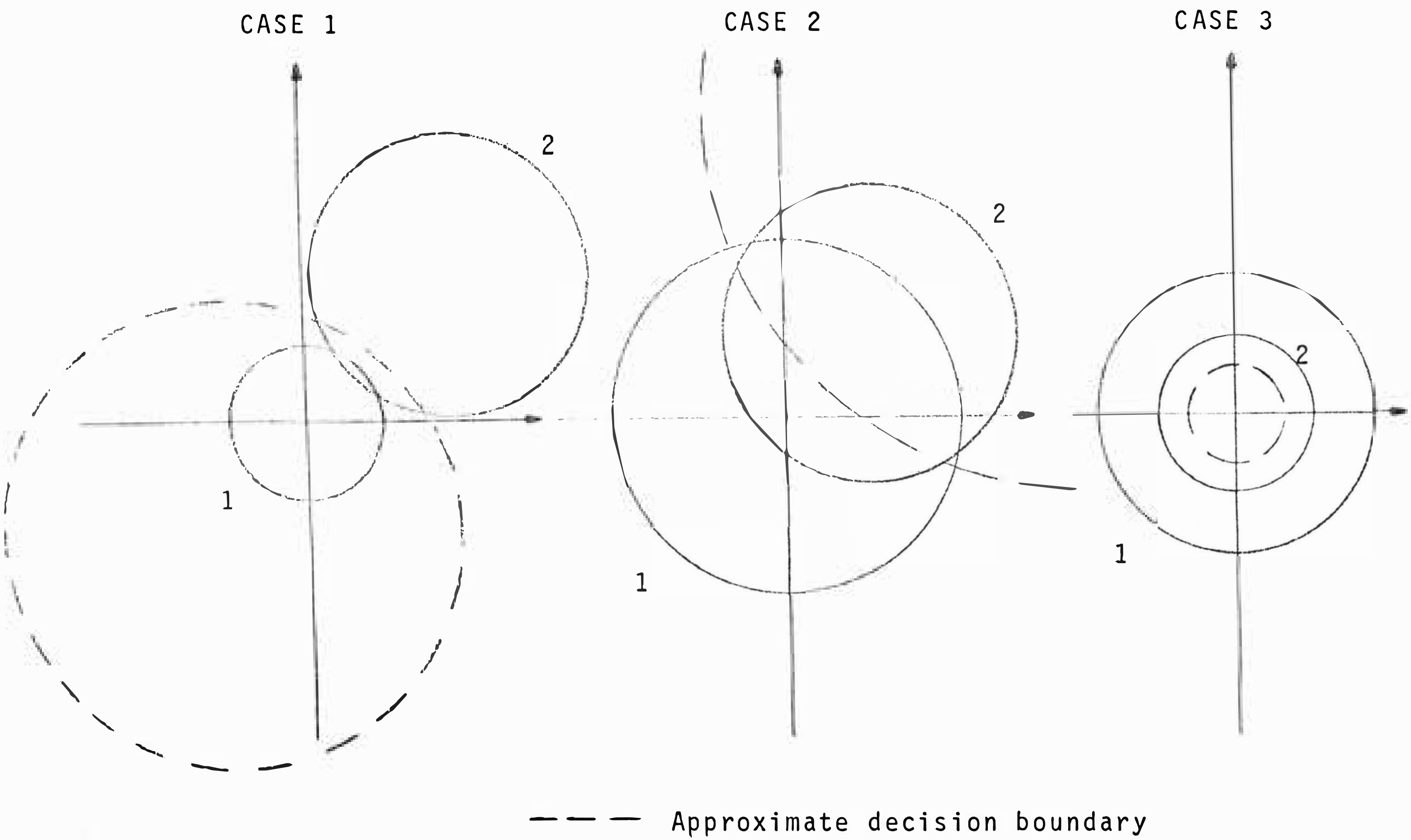


Figure 5.3. Three-sigma contours of constant probability density for two classes of distributions.

TABLE 5.1				
CONFUSION MATRIX-A PRIORI GAUSSIAN PDF BAYES' RECEIVER				
Case	Signal		S1	S2
1	GAUSS	S1	125	0
		S2	1	124
	CAUCHY	S1	101	24
		S2	3	122
2	GAUSS	S1	107	18
		S2	13	112
	CAUCHY	S1	105	20
		S2	30	95
3	GAUSS	S1	78	47
		S2	16	109
	CAUCHY	S1	86	39
		S2	57	68

important fact that the transformation must be one-to-one; i.e., that for each point in the original space A there corresponds one, and only one, point in B; and, that to each point in B there corresponds one, and only one, point in A. In the prior definition A denotes the space $\{x|x$ is the set of values for which the probability density function $p(x)$ is defined$\}$, and the space B is $\{y|y$ is the set of values defined by the transformation $y=f(x)\}$.

For example, Hogg and Craig [23] work, as an example, the following special problem. Let x be a continuous random variable, having a probability density function $p(x)=2x$, $0<x<1$. A is the space $\{x|0<x<1\}$. Under the transformation $y=8x^3$, space A maps into $B=\{y|0<y<8\}$. The transformation is one-to-one.

The authors Hogg and Craig discuss multivariate transformations as well, their conclusions very briefly summarized here. Let the functions $y_i=f_i(x_1,x_2,\ldots,x_n)$ and the inverse functions $x_i=g_i(y_1,y_2,\ldots,y_n)$, $i=1,2,\ldots,n$ define a one-to-one transformation which maps A into B. Form the Jacobian as the n x n determinant of the partial derivatives of the inverse functions

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} .$$

The joint probability density function of the random variates is given by $P_Y(y_1, y_2, \ldots, y_n) = |J| P_X[g_1(y_1, y_2, \ldots, y_n), \ldots, g_n(y_1, y_2, \ldots, y_n)]$. If the selected transformation is not one-to-one, the difficulty of the problem is multiplied k-fold, where k represents the number of mutually exclusive events required to define transformations which are one-to-one over their range. Instead of writing $|J|$, one now writes $|J_i|$; and $x_n = g_n(y_1, y_2, \ldots, y_n)$ is replaced by $x_{ij} = g_{ij}(y_1, y_2, \ldots, y_n)$, for $j = 1, 2, \ldots, k$.

The complexity of the one-to-one transformation and the tedious computations required to solve for the joint probability density function serve to make the procedure less than appealing. Some transformations were proposed, and the Jacobian evaluated, but several of the resulting expressions for alpha and beta required numerical integration which was not feasible. Without an evaluation of alpha and beta the goodness of the transformation cannot be evaluated. For these reasons further investigation was suspended.

III. Discussion of Transformation $z=x^2$

Originally, a transformation of variates by $z=x^2$ was chosen for evaluation based on a hypothesis that one can square Gaussian variates of mean μ and unit variance, and can classify the resulting data, having a chi-squared probability density function, with fewer errors than the original data. The optimum receiver is Bayes' receiver designed for a priori chi-squared variates.

If $\chi^2=x_1{}^2+x_2{}^2+\ldots+x_v{}^2$, where $x_1,x_2,\ldots,x_v$ are independent random variates Gaussianly distributed with zero mean and unit variance, then the probability density function of χ is given by

$$p(\chi^2) = \left\{\frac{1}{2^{\frac{v}{2}}\,\Gamma(v/2)}\,(\chi^2)^{\left(\frac{v}{2}-1\right)}\,e^{\frac{-\chi^2}{2}}\right\} \tag{8}$$

where v represents the degrees of freedom, here meaning the number of independent random variates. For a Gaussian random variate of mean μ and variance σ^2, the random variable

$$z = \left(\frac{x-\mu}{\sigma}\right)^2$$

has a χ^2 distribution with one degree of freedom; the probability density function is an exponential curve.

Because the variance is constrained to unity, the problem of two classes of data requires that the mean of one distribution be non-zero. The non-central chi-squared distribution is a generalization of the previous distribution when the mean is non-zero. One class of data is described by the following probability density function of z, given by Anderson [24], where $z = x^2$.

$$p(z) = \left\{ \frac{1}{2^{\frac{v}{2}}\sqrt{\pi}} e^{-\frac{1}{2}(w^2+z)} z^{(\frac{w}{2}-1)} \sum_{n=o}^{\infty} \frac{(w^2)^n z^n \Gamma(n+\frac{1}{2})}{(2n)!\,\Gamma(\frac{1}{2}v+n)} \right\}, \qquad (9)$$

where w equals the mean squared.

Equation (9) cannot be solved for z explicitly, and for this reason further study was discontinued. Lacking an expression $z = f(v,w)$ which can be compared to a threshold, the transformation is not useful in this research effort. Although one could substitute each data vector [z] into the above equation and obtain a number to compare to a threshold, a method for evaluation of the resulting classification errors is not available; as a result, a numerical value for alpha and beta cannot be obtained. Also, experience with a series expansion technique to calculate the error function, discussed in Appendix 1, indicates that between fifty and seventy-five terms may be required for the sum to converge. In the process, z^n or $(2n)!$ may exceed the computer register, which for the IBM 360 model 30 used by the author, is 10^{75} and 10^{-78}. Figure 5.4 illustrates the χ^2 distribution.

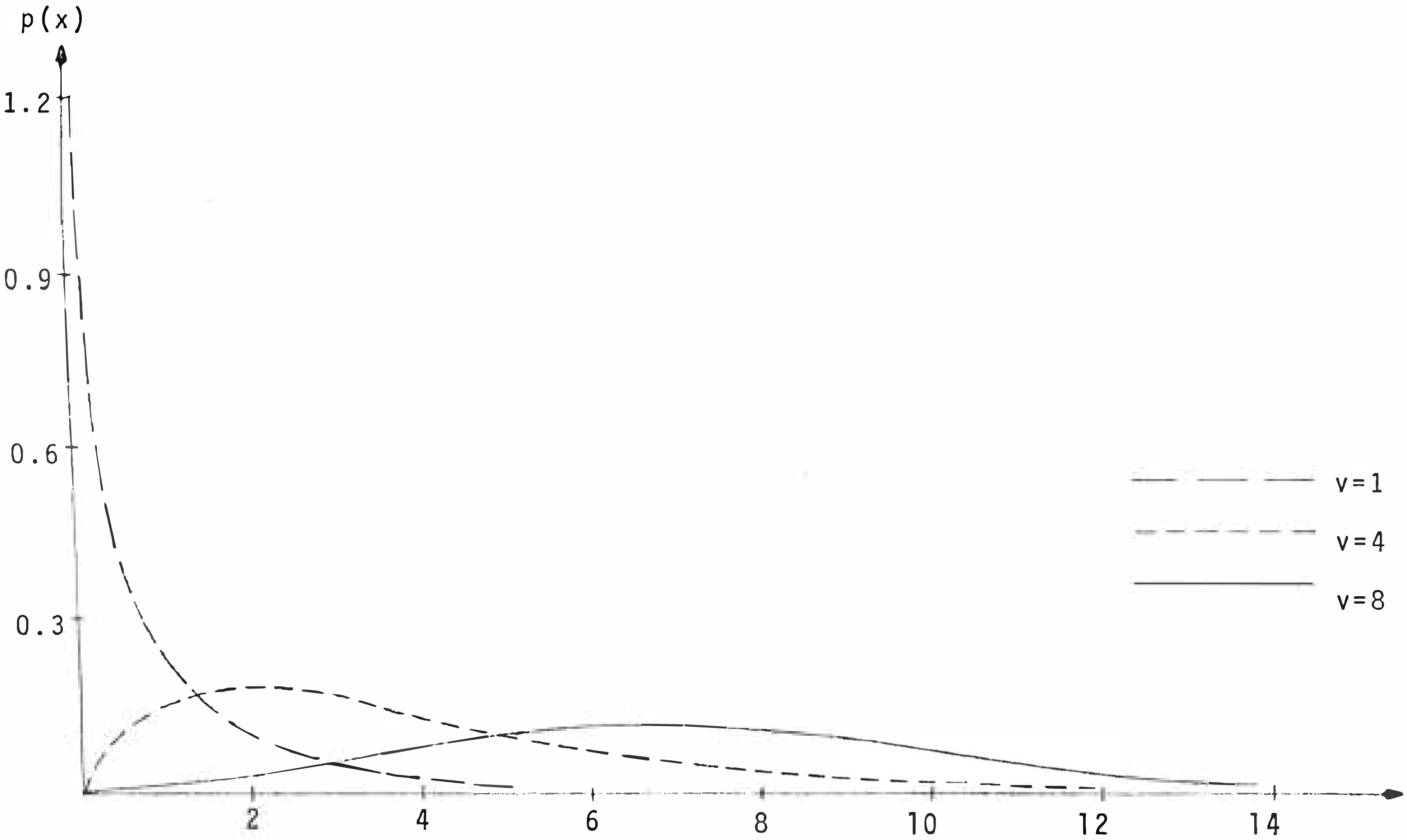


Figure 5.4. χ^2 probability density function for selected values of v.

IV. The Weibull Distribution

Consideration was given to classification of variates by a Bayes' receiver designed to classify Weibull distributed variates. Depending upon the shape factor s, the distribution can be made to look like several commonly encountered distributions. Two other parameters of the distribution, the scale factor h and the location factor l, are required to complete the description of the Weibull distribution. Grant [25] was the source for the probability density function, which is written as

$$p(x) = \left\{ \frac{s}{h} \left(\frac{x-l}{h}\right)^{s-1} \exp\left[-\left(\frac{x-l}{h}\right)^{s}\right] \right\} \text{ for } \begin{matrix} s>0.0 \\ h>0.0 \\ x>l \end{matrix} \quad . \tag{10}$$

Commonly l=0 and h=1.

For s = 1, the equation describes the exponential probability density function; when s = 2, the equation describes a Rayleigh probability density function; and for $3.5<s<4.0$ the equation is an approximation to the Gaussian probability density function. This means that one receiver written for a Weibull distribution could classify exponential, Rayleigh, and Gaussian distributed variates simply by changing the shape factor s. Figure 5.5 illustrates the Weibull distribution for three values of shape factor. Complications arise, however when the distribution is to be used in a classification problem. As is true with some other distributions studied, the decision boundary cannot

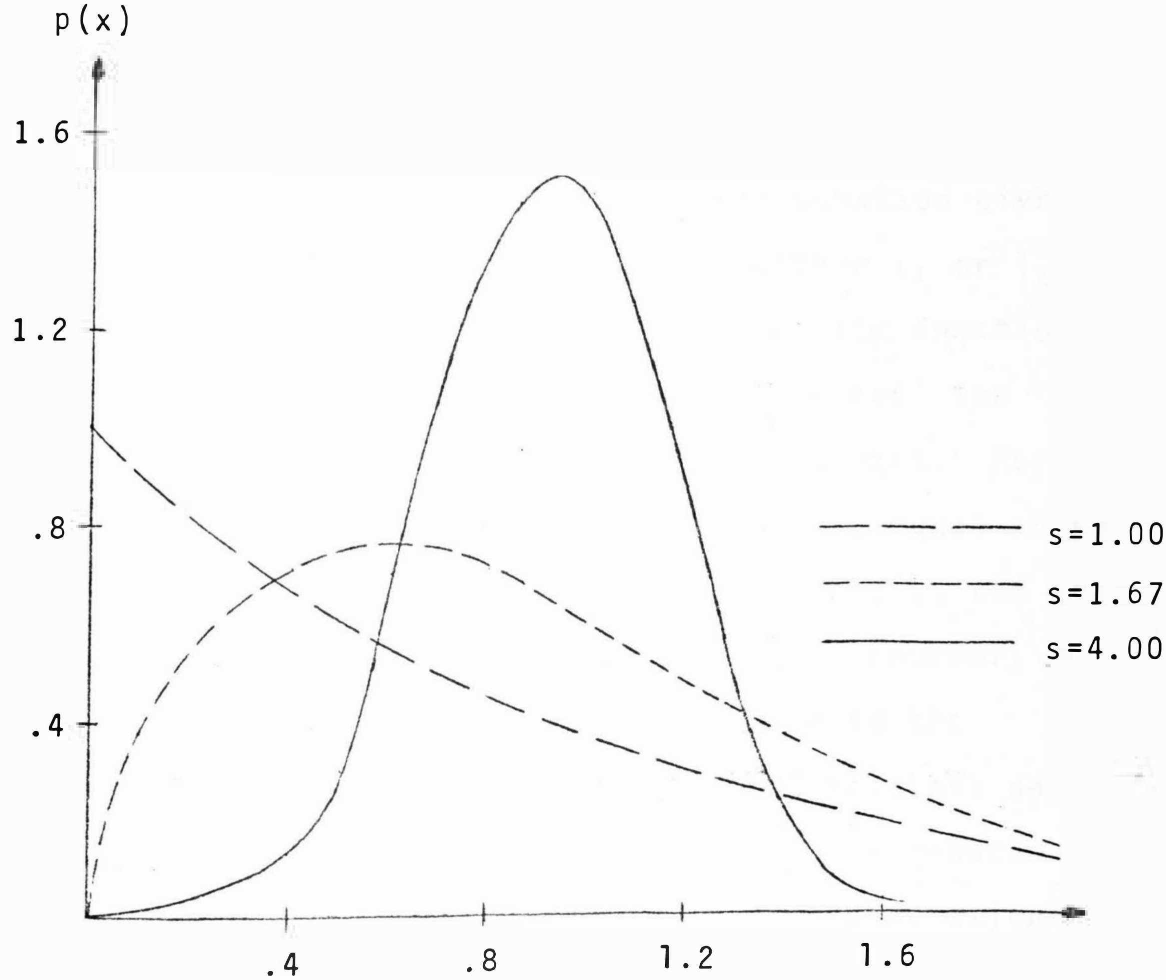


Figure 5.5. Weibull probability density function for selected values of shape factor s.

be written with x as a function only of K and the shape factors or location factors used to describe the two distributions. As a consequence of this limitation, the decision boundary cannot compare the sample data vector with a threshold value, but must solve an equation in x.

A second disadvantage arises because of a natural constraint on the allowable value of x. The random variate x cannot be used in the decision boundary equation given in Table 2.1 if its magnitude is less than either l_1 or l_2, because (10) ceases to be a probability density function for x less than l. In a practical two class decision problem the above criterion is impossible to meet. For example, consider two distributions which have equal shape factors. One class is located at l_1 equal to 0.0, the other has l_2 equal to 2.0. Logically, the decision boundary must be located somewhere between l_1 and l_2. Due to the constraint imposed on the random variate x, that all sample data points be greater than 2.0, the problem is reduced to one class, and no decision is required. The conclusion must be made that the Weibull distribution is of little practical use for this classification problem.

CHAPTER 6: CONCLUSIONS

This final chapter seems to the author as much a preface to future research by others as it is conclusions of this thesis. Although much time and considerable effort has been expended in search of sometimes elusive answers, rewards were reaped two-fold. First, the personal store of knowledge was increased as a direct result of the research, and secondly, questions which merit further study by others were raised. The combination is hardly unique, but most certainly beneficial.

The primary objective, evaluation of the transformation $z = y_1/y_2$, was achieved. Contrary to the belief held at the beginning of the research effort, the transformation neither yields an improvement in alpha and beta error, nor qualifies as an approximation for the alpha and beta error of the original distribution. For the two examples evaluated, Table 6.1 reviews the results of the transformation on alpha error, beta error, and total probability of error. The author believes that if the problem were repeated using different values of variance and correlation coefficient, the original data would still yield the lower classification error. This important conclusion implies that fewest number of alpha and beta errors are made if the original (non-transformed) random variates are classified.

Example	ρ	Distribution	α	β	Total Error
1	0.90	Bivariate Gaussian	0.370	0.122	.246
		Univariate Cauchy	0.392	0.216	.304
2	0.98	Bivariate Gaussian	0.216	0.053	.1345
		Univariate Cauchy	0.347	0.130	.2385

Table 6.1. Classification error for original and transformed distributions.

Some dependency of the classification error on the correlation coefficient may be noted from Table 6.1. As the correlation coefficient increases toward unity, the total probability of error for the receiver designed to classify the original Gaussian random variates decreases. The total probability of error for the receiver designed to classify the transformed Cauchy variate decreases also, but by a proportionally smaller amount. For zero correlation of variates, the two receivers have exactly equal classification error, alpha equal to zero and beta equal to unity if the threshold K is greater than unity; i.e., the receiver biased in favor of class one. If class K is less than unity, alpha is equal to unity and beta is equal to zero; i.e., the receiver is biased in favor of class two. Alpha and beta equal 0.5 for K equal to unity. The author suggests that for any value of correlation coefficient greater than zero, the receiver for the original distribution of data has the lower classification error.

One may argue that a lower classification error cannot be obtained with any transformation which involves a reduction of dimensionality, such as the transformation $z=y_1/y_2$. The desirability of a multi- rather than uni-dimensional data vector for classification may be illustrated by saying each dimension is an observation. Generally speaking, several observations are preferable to only one. In the above transformation, a 2-dimensional data vector is traded for a 1-dimensional data vector; this loss in dimensionality represents a loss of information. Depending on the actual value of the data vector, the loss may or may not be significant. However, for a large number of data vectors (the alpha and beta error assume an infinite number of vectors) some transformations do represent a significant loss of information; consequently, the total probability of error for the transformed data vectors must be greater than the total probability of error for the original data vectors. If this argument is true, then the results obtained from this research, summarized in Table 6.1, are in agreement with the proposed theory. However, no proof is offered.

The other areas of study, which were discussed in Chapter 5, were summarized in their respective sections, and will receive only cursory comments here.

The greatest disadvantage of the receiver optimal for the non-central chi-squared distribution is that the decision boundary equation cannot be expressed explicitly in terms of the variable χ^2. If a good approximation to the infinite sum contained in the chi-squared probability density function were formulated, the transformation $\chi^2 = x_1^2 + x_2^2 + \ldots + x_v^2$ may merit further study. The transformation provides an opportunity to verify the argument that a reduction in dimensionality is detrimental to the goal of low classification error.

Many man-months could be devoted to the selection and evaluation of bivariate and multivariate transformations. One must be prepared to invest considerable amounts of both his own time and computer time for the evaluation of integrals, inverse transformations, and determinants, however.

Although it has application in quality control, the Weibull distribution was prevented from being useful to this research because of the requirement that the data vector always be larger than either of the two location parameters l_1 and l_2. Future investigators should take note of this limitation.

More thought could be given to an unexpected result obtained from the Cauchy-Gauss simulation discussed in section one of Chapter 5. The decision boundary in Figure 5.3(a) and 5.3(b) is a closed loop, although it appears that fewer errors would be made if the decision boundary were instead a curved infinite line. In Figure 5.3(c) the decision boundary is a circle, which is the expected optimum boundary.

This completes the discussion of research results and recommendations for areas of future study. A discussion of techniques for computation of the error function is included in Appendix 1. A derivation of the Cauchy distribution from the bivariate Gaussian distribution is included in Appendix

2. The two main computer programs used in this research are reproduced in Appendix 3.

REFERENCES

[1] J. C. Hancock and P. A. Wintz, Signal Detection Theory. New York: McGraw-Hill, 1966, pp. 30-59.

[2] J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering. New York: Wiley, 1965, pp. 37-84.

[3] W. B. Davenport Jr. and W. L. Root, An Introduction to the Theory of Random Signals and Noise. New York: McGraw-Hill, 1958, pp. 32-38, 145-154, 251-254.

[4] A. Papoulis, Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill, 1965, pp. 187-199.

[5] K. Fukunaga and T. F. Krile, "Calculation of Bayes' Recognition Error for Two Multivariate Gaussian Distributions," IEEE Transactions on Computers, Vol. C-18, No. 3, pp. 220-229, March 1969.

[6] T. W. Anderson, An Introduction to Multivariate Statistical Analysis. New York: Wiley, 1958, pp. 10, 112.

[7] R. V. Hogg and A. T. Craig, Introduction to Mathematical Statistics. New York: Macmillan, 1959, pp. 31-39, 78-84.

[8] N. L. Johnson and F. C. Leone, Statistics and Experimental Design in Engineering and the Physical Sciences, Vol. 1. New York: Wiley, 1964, pp. 100-139.

[9] T. H. Naylor, et. al., Computer Simulation Techniques. New York: Wiley, 1966, pp. 77-118.

[10] F. B. Hildebrand, Introduction to Numerical Analysis. New York: McGraw-Hill, 1956, pp. 71-82.

[11] M. Abramowitz and I. S. Stegun, Eds., Handbook of Mathematical Functions. New York: Dover Publications, 1965, pp. 297-311.

[12] N. M. Blachman, Noise and Its Effect on Communication. New York: McGraw-Hill, 1966, pp. 3-16.

[13] W. G. Wee, "A Survey of Pattern Recognition," presented at the 7th IEEE Symposium on Adaptive Processes, 1968.

[14] J. C. Hancock and P. A. Wintz, pp. 32-45.

[15] N. M. Blachman, p. 7.

[16] W. B. Davenport and W. L. Root, pp. 147-148.

[17] F. F. Martin, Computer Modeling and Simulation. New York: Wiley, 1968, p. 59.

[18] R. V. Hogg and A. T. Craig, p. 73.

[19] J. C. Hancock and P. A. Wintz, p. 44.

[20] T. H. Naylor, et. al., pp. 97-99.

[21] J. M. Wozencraft and I. M. Jacobs, pp. 58-63.

[22] J. M. Wozencraft and I. M. Jacobs, pp. 111-114.

[23] R. V. Hogg and A. T. Craig, pp. 65-66.

[24] T. W. Anderson, pp. 112-113.

[25] E. L. Grant, Statistical Quality Control. New York: McGraw-Hill, 1964, p. 506.

[26] M. Abramowitz and I. S. Stegun, p. 297.

[27] D. D. McCracken, Fortran with Engineering Applications. New York: Wiley, 1967, pp. 60-72.

[28] M. Abramowitz and I. S. Stegun, p. 299.

[29] C. Hastings Jr., Approximations for Digital Computers. Princeton, N. J.: Princeton University Press, 1955.

APPENDIXES

APPENDIX 1: COMPUTATION OF THE ERROR FUNCTION

Both the expression for alpha and the one for beta contained an integral of the product of the error function times an exponential function. Rather than tabulate the error function for several hundred values of the integration variable, the error function was computed as an integral part of the computer program. Initially, the author used the series expansion definition of the error function as written in Abramowitz and Stegun [26],

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!\,(2n+1)}$$

The series was found to converge for as few as ten terms for small (<0.5) values of the argument. The error function was accurate to the sixth decimal for arguments to 1.5, and accurate to at least the fourth decimal for arguments from 1.5 to 2.0. Accuracy was not checked for arguments greater than 2.0.

Inserted into the main program, the error function calculation produced the same fine results until the error function argument increased to approximately 3.0. At that value exponent overflow error in the computer register occurred; the term (x^{2n+1}) exceeded 10^{75}, and the denominator term of $[n!\,(2n+1)]$ was approaching the register limit.

Previously it had been found that if the denominator term were expressed in fixed point notation, the computer register limit of $(2^{32}-1)$ was exceeded on the twelfth term of the summation. The overflow error had not been anticipated in floating point notation, however.

A scale factor was introduced in both numerator and denominator to reduce the magnitude of the offending terms. A scale factor of 10^{-62} was chosen to avoid underflow error and to permit computation of arguments as large as 3.5 before overflow error occurred. After the addition of the scale factor the sum was found to diverge when the error function argument exceeded 3.5. No explanation seems reasonable except round-off error of the denominator term n!. (For n=75, n! is approximately $(8. \times 10^{70})$, using Sterling's approximation to the factorial). Making all variables double precision real permitted calculation of the error function for arguments as large as 5.5 before the summation began to diverge. Observation of the magnitude of n! and the improvement resulting from use of double precision definitely makes round-off error appear responsible for the divergence of the summation. For most common applications, the error function of arguments (>5.0) can be assumed to be equal to 1.00, and computation of large arguments then becomes unnecessary. For a more complete discussion on the subject of errors due to the limitation of digit storage in computer memory, read McCracken [27].

Equations which approximate the error function are contained in Abramowitz and Stegun [28] which were original with Hastings [29]. Equation 7.1.26,

$$erf(x) = 1.0 - (a_1t + a_2t^2 + a_3t^3 + a_4t^4 + a_5t^5)e^{-x^2}$$

$$\text{where } t = \frac{1}{1 + px},$$

was programmed, and the results compared to the series expansion. The rational approximation, which is equation 7.1.26, was as accurate as the series expansion.

The rational approximation method has an advantage over the series expansion method of error function computation, and that advantage is significantly reduced computation time. One calculation instead of up to fifty is required for each argument. In addition, error functions for arguments greater than 5.0 may be computed.

Equation 7.1.26 is valid only for positive x, however, so one must make the argument positive via an IF statement in the program. The symmetry property of the error function is invoked, namely, erf(-x) = erf(x). The speed and accuracy, as well as the freedom from round-off error and register overflow, make the rational approximation method far preferable to the series expansion technique.

APPENDIX 2: DERIVATION OF CAUCHY DISTRIBUTION FROM BIVARIATE GAUSSIAN DISTRIBUTION

If y_1 and y_2 are jointly normal with

$$f(y_1,y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left[\frac{-1}{2(1-r^2)}\left(\frac{y_1{}^2}{\sigma_1{}^2}-\frac{2ry_1y_2}{\sigma_1\sigma_2}+\frac{y_2{}^2}{\sigma_2{}^2}\right)\right] \tag{1}$$

then the density function of the random variable $z = y_1/y_2$ is developed briefly in the following steps.

$$f(z) = \frac{2}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^\infty y_2 \exp\left[\frac{-y_2{}^2}{2(1-r^2)}\left(\frac{z^2}{\sigma_1{}^2}-\frac{2rz}{\sigma_1\sigma_2}+\frac{1}{\sigma_2{}^2}\right)\right] dy_2 \tag{2}$$

is obtained using the relation developed in Papoulis

$$f(z) = 2 \int_0^\infty y_2 \, f_{y_1y_2}(zy_2,y_2)\, dy_2 \qquad \text{if} \tag{3}$$

$f(y_1,y_2) = f(-y_1,-y_2)$, i.e. has symmetrical masses. To obtain the solution to the above integral one must use the relation

$$\int_0^\infty y \exp \frac{-y^2}{2a^2}\, dy = a^2 \int_0^\infty e^{-w} dw = a^2 \tag{4}$$

which is obtained using the substitution of variables $w = y^2/2a^2$, $dy = a^2 y dw$.

$$f(z) = \left(\frac{1}{\pi\sigma_1\sigma_2\sqrt{1-r^2}}\right) \quad (1-r^2) \quad \left(\frac{1}{\frac{z^2}{\sigma_1{}^2} - \frac{2rz}{\sigma_1\sigma_2} + \frac{1}{\sigma_2{}^2}}\right) \tag{5}$$

$$f(z) = \frac{\sqrt{1-r^2} \quad \sigma_1\sigma_2}{\pi \quad \left[\sigma_2{}^2 \left(z^2 - 2rz\frac{\sigma_1}{\sigma_2}\right) + \sigma_1{}^2\right]} \tag{6}$$

By completing the square on the first denominator term one obtains the desired result

$$f(z) = \frac{\sqrt{1-r^2} \quad \sigma_1\sigma_2}{\pi \quad \left[\sigma_2{}^2 \quad \left(z - r\frac{\sigma_1}{\sigma_2}\right)^2 + \sigma_1{}^2(1-r^2)\right]} \tag{7}$$

If the jointly Gaussian variates are uncorrelated,

$$f(z) = \frac{\sigma_1}{\pi\sigma_2 \quad (z^2 + \sigma_1{}^2/\sigma_2{}^2)} \; . \tag{8}$$

Justification for (3) may be obtained from Papoulis [13], who was the source for this transformation derivation. It should be noted, however, that his integral equation similar to (4) contains an incorrect intermediate equation, although he concludes with the correct result.

```
C    COMPUTATION OF ERROR OF THE FIRST KIND, ALPHA, AND ERROR OF THE SECOND KIND
C    BETA, BY NUMERICAL INTEGRATION USING THE TRAPEZOIDAL RULE.  THE REQUIRED
C    ERROR FUNCTION IS COMPUTED TO AN APPROXIMATE VALUE BY THE RATIONAL
C    APPROXIMATION METHOD OUTLINED BY HASTINGS, APPROXIMATIONS TO DIGITAL
C    COMPUTERS, 1956.  THE CLASS 2 DATA WHICH IS CORRELATED WITH RHO EQUAL TO 0.9
C    IS UNCORRELATED USING THE ROTATION OF COORDINATES OUTLINED IN DAVENPORT AND
C    ROOT TO EFFECT THE DESIRED TRANSFORMATION.
C
  201 FORMAT ('1','            ALPHA ERROR =',F10.6)
  202 FORMAT ('0',50X,'INDIVIDUAL TERMS OF G(J)')
  203 FORMAT ('0',10F11.5)
  206 FORMAT ('1','             BETA ERROR =',F10.6)
C
C    DEFINITION OF INITIAL VALUES,UPPER LIMIT,INCREMENT OF Y,AND COMPUTATION
C    OF CONSTANTS
C
      DOUBLE PRECISION ERF
      DIMENSION F(840),G(840)
C
      R=0.90
      A=1.0/(2.0*(1.0&R))
      B=SQRT(2.0*(1.0-R))
      C=B/SQRT(3.1416*(1.0-R*R))
      D=SQRT(2.0/3.1416)
C
C    DO LOOP TO COMPUTE ALPHA AND BETA FOR TWO CLASSIFICATION PROBLEM SETS
C
      DO 3 K=1,6
C
      GO TO (20,21,22,50,51,52),K
   20 I=1
      LIMU=400
      DELTX1=0.01
      GO TO 23
   21 I=1
      LIMU=520
```

```
      DELTX1=0.005
      GO TC 23
   50 R=0.98
      A=1.C/(2.C*(1.0&R))
      B=SQRT(2.C*(1.0-R))
      C=B/SQRT(3.1416*(1.0-R*R))
      I=I
      LIMU=400
      DELTX1=0.C1
      GO TC 23
   51 I=1
      LIMU=460
      DELTX1=0.005
   23 G(1)=0.0
      X1=0.0
      GO TC 24
   22 I=521
      LIMU=800

      GO TO 24
   52 I=461
      LIMU=838
   24 CONTINUE
C
C     DO LCOP TO COMPUTE VCLUME BY TRAPEZOIDAL RULE
C
      DO 2 J=I,LIMU
C
      GC TC (11,12,13,53,54,55),K
   11 TAU=(1./1.414)*SQRT(0.1845+0.0526*X1*X1)
      GO TO 14
   12 TAU=(1./B)*SQRT(0.1845+0.0526*X1*X1)
      GO TC 14
   13 TAU=(1./B)*SQRT(0.90-0.0526*X1*X1)
      GO TC 14
   53 TAU=(1./1.414)*SQRT(0.066+0.0101*X1*X1)
      GO TC 14
```

```
   54 TAU=(1./B)*SQRT(0.066+0.01C1*X1*X1)
      GO TO 14
   55 TAU=(1./B)*SQRT(0.180-0.0101*X1*X1)
C
C    ERROR FUNCTION CCMPUTATICN BY RATIONAL APPROXIMATION WITH ERROR TERM 10**-7
C
   14 IF (TAU)5,6,6
    5 TAUNEG=-1.0*TAU
      T=1.0/(1.C&0.327591*TAUNEG)
      GO TC 7
    6 T=1.0/(1.C&0.327591*TAU)
    7 TSQRD=T*T
      TCUBED=T*TSQRD
      TQUAD=T*TCUBED
      TPENT=T*TQUAD
C
      ERF=1.0-(0.25482959D0*T-0.28449674D0*TSQRD&1.42141374D0*TCUBED-1.4
     15315203D0*TQUAC&1.06140543C0*TPENT)*EXP(-(TAU*TAU))
      IF (TAU)8,9,9
    8 ERF=-1.0*ERF
    9 GO TC (15,16,16,15,16,16),K
C
C  CCMPUTATICN OF THE INTEGRANC F(X)
C
   15 F(J)=ERF*EXP(-1.0*X1*X1*0.5)
      GC TC 17
   16 F(J)=ERF*EXP(-1.0*X1*X1*A)
   17 CONTINUE
      X1=X1&DELTX1
C
    2 CONTINUE
C
C    CCMPUTATION OF THE INTEGRAL BY SIMPSON'S RULE
C
      CALL QSF (DELTX1,F,G,LIMU)
```

```
      GO TO (18,3,19,18,3,19),K
   18 ALPHA=G(LIMU)*C
      WRITE (12,201) ALPHA
      GO TO 1
   19 BETA=1.0-G(LIMU)*C
      WRITE (12,206) BETA
      GO TO 1
    1 WRITE (12,202)
      WRITE (12,203) (G(J),J=1,LIMU)
C
    3 CONTINUE
C
      END
```

```
C      THIS PROGRAM IS LABELED CAUCHY-GAUSS-RECEIVER.  UNIVARIATE
C      CAUCHY DATA IS GENERATED BY THE TRANSFORMATION Z=X1/X2, WHERE X1,
C      X2 ARE GAUSSIAN VARIATES OF VARIANCE SIGMA1, SIGMA2 AND MEAN ZERO.
C      BIVARIATE GAUSSIAN DATA OF VARIANCE SIG1, SIG2 AND MEAN 0, MEAN A
C      IS GENERATED USING SUBROUTINE GAUSS. BOTH SETS OF SIGNALS ARE
C      RECEIVED AND CLASSIFIED BY A GAUSSIAN OPTIMAL BAYES RECEIVER.
C      REFERENCES FOR THIS WORK ARE PAPOULIS (PROBABILITY, RANDOM
C      VARIABLES, AND STOCHASTIC PROCESSES), PP 198-200 AND HANCOCK,WINTZ
C      (SIGNAL DETECTION THEORY), PP. 30-86.
C
C
C      GAUSSIAN OPTIMIZED BAYES' RECEIVER FOR CLASSIFICATION OF CUACHY R.V.
C
C
       DIMENSION U(10),X(10),Y(10)
       REAL MEANA
C
C      INPUT DATA READ
C
    23 READ (11,102) SIG1,SIG2,MEANA,COSTK
       WRITE (12,212) SIG1,SIG2,MEANA,COSTK
       READ (11,101) IX,SS,AM
       WRITE (12,211) IX,SS,AM
       READ (11,100) R,SIGMA1, SIGMA2
C
C      COMPUTATION OF CONSTANTS
C
       A=SIG1*SIG1
       B=SIG2*SIG2
       C=B-A
       D=2.0*MEANA*A
       E=D*MEANA
       F=2.0*A*B*ALOG((B*COSTK)/A)
       ALPHA=0.0
       BETA=0.0
       CL1OK=0.0
```

```
      CL2OK=0.0
      WRITE (12,210) R,SIGMA1,SIGMA2
      WRITE (12,207)
C
C     DO LCOP TO CONTROL THE TYPE OF DISTRIBUTION TO BE GENERATED.
C
      DO3 K=1,4
      MM=K
C
C     DO LOOP TO GENERATE 125 TWO-DIMENSIONAL RANDOM VARIATES
C
      DO 2 J=1,125
      GO TO (10,10,11,11), MM
C
C     GENERATION OF CAUCHY VARIABLES
C
   10 DO1 I=1,2
      CALL GAUSS(IX,SS,AM,V)

      X(I)=SIGMA1*V
      TS=R*SIGMA2*V
      CALL GAUSS(IX,SS,AM,V)
      Y(I)=TS&(SQRT(1.0-R*R)*SIGMA2*V)
      U(I)=X(I)/Y(I)
    1 CONTINUE
      WRITE (12,200) (U(I),I=1,2)
      GO TO 21
C
C     GENERATION OF GAUSSIAN VARIABLES
C
   11 DO4 L=1,2
      CALL GAUSS(IX,SS,AM,V)
      U(L)=V
    4 CONTINUE
      WRITE (12.219) (U(L).L=1,2)
```

```
C      CLASSIFICATION OF CAUCHY AND GAUSSIAN VARIATES BY (GAUSSIAN) BAYES
C      RECEIVER
C
   21 G=C*U(1)*U(1)&C*U(2)*U(2)&C*(U(1)&U(2))-E-F
C
C      DECISION ANNOUNCED AND ERRORS IN CLASSIFICATION RECORDED
C
      GO TO (12,13,12,13), MM
   12 IF(G)14,14,15
   13 IF(G)16,16,17
   14 WRITE (12,201)
      CL1OK=CL1OK&1.0
      GO TO 2
   15 WRITE (12,202)
      ALPHA=ALPHA&1.0
      GO TO 2
   16 WRITE (12,203)
      BETA=BETA&1.0
      GO TO 2
   17 WRITE (12,204)
      CL2OK=CL2OK&1.0
    2 CONTINUE
      GO TO (18,19,20,22), MM
C
C      TOTAL ERRORS FOR ONE CLASS RECORDED AND DATA FOR SECOND CLASS READ
C
   18 READ (11,100) R,SIGMA1,SIGMA2
      WRITE (12,210) R,SIGMA1,SIGMA2
      GO TO 3
   19 WRITE (12,218)
      WRITE (12,215)
      WRITE (12,216) CL1OK,ALPHA
      WRITE (12,217) BETA,CL2OK
C
C      DATA FOR GAUSSIAN VARIABLES READ IN
C
      ALPHA=0.0
```

```
      BETA=0.0
      CL1OK=0.0
      CL2OK=0.0
      WRITE (12,208)
   20 READ (11,101) IX,SS,AM
      WRITE (12,211) IX,SS,AM
    3 CONTINUE
   22 WRITE (12,218)
      WRITE (12,215)
      WRITE (12,216) CL1OK,ALPHA
      WRITE (12,217) BETA,CL2OK
C
C     RETURN FOR GENERATION AND CLASSIFICATION OF NEXT PROBLEM SET
C
      GO TO 23
  100 FORMAT (3F10.4)
  101 FORMAT (I5,2F10.4)
  102 FORMAT (4F10.4)
  200 FORMAT (90X,F7.3,11X,F7.3)
  201 FORMAT (1H+,'SIGNAL 1 CLASSIFIED AS SIGNAL 1, CORRECT DECISION')
  202 FORMAT (1H+,'SIGNAL 1 CLASSIFIED AS SIGNAL 2, ERROR OF THE
     1FIRST KIND')
  203 FORMAT (1H+,'SIGNAL 2 CLASSIFIED AS SIGNAL 1, ERROR OF THE
     1SECOND KIND')
  204 FORMAT (1H+,'SIGNAL 2 CLASSIFIED AS SIGNAL 2, CORRECT DECISION')
  207 FORMAT ('0','CLASSIFICATION OF CAUCHY VARIATES,SIGNALS ONE AND TWO
     1',37X,'CAUCHY VARIATES ARE'/91X,'U1=',14X,'U2=')
  208 FORMAT ('1','CLASSIFICATION OF GAUSSIAN VARIATES, SIGNALS ONE AND
     1TWO',35X,'GAUSSIAN VARIATES ARE'/89X,'U1=',14X,'U2=')
  210 FORMAT ('0','R=',F8.4,'   SIGMA1=',F8.4,'   SIGMA2=',F8.4)
  211 FORMAT ('0','IX=',I6,'   SS=',F8.4,'   AM=',F8.4)
  212 FORMAT ('1','SIG1=',F8.4,'   SIG2=',F8.4,'   MEANA=',F8.4,'   COSTK
     1=',F8.4)
  215 FORMAT ('0',28X,'S1',8X,'S2')
  216 FORMAT ('0',24X,'S1',1X,F6.1,4X,F6.1)
  217 FORMAT ('0',24X,'S2',1X,F6.1,4X,F6.1)
  218 FORMAT ('0',26X,'CONFUSION MATRIX')
  219 FORMAT (88X,F7.3,9X,F7.3)
      END
```

```
C    COMPUTATION OF ERROR OF THE SECOND KIND, BETA , USING NUMERICAL INTEGRATION
C    BY THE TRAPEZOIDAL RULE.
C
  200 FORMAT ('0','  CUMULATIVE ERFAY       CUMULATIVE ERFBU    SUM TO
     1AFYB**2N&1      FACTN1*2N&1       BSQA**2N&1       FACTN2*2N&1')
  201 FORMAT (4X,D14.6,8X,D14.6,7X,I3,5X,D10.3,6X,D10.3,7X,D10.3,6X,D10.
     13)
  202 FORMAT ('0',4X,'O= ',E10.3,'    P= ',E10.3,'    SCALE1= ',D10.3,'
     1SCALE2= ',D10.3,'    LIMITU= ',I3,'    DELTAY= ',D10.3)
  205 FORMAT ('0','ARGUMENT Y      COMPUTED F(Y)      ERF ARGUMENT AFYB
     1  ERROR FUNCTION ERFAY      ERF ARGUMENT BSQA      ERROR FUNCTION ER
     1FBY')
  206 FORMAT (3X,F5.2,10X,E9.2,11X,F10.6,10X,D17.10,10X,F10.6,9X,D17.10)
  207 FORMAT ('1','BETA ERROR= ',F10.5,24X,'INDIVIDUAL TERMS OF G(J)=')
  208 FORMAT ('0',35X,5F10.5)
C
      DOUBLE PRECISION R,A,SQA,CB,CC,D,E,Y,B,C,ABC,ULIMIT,AFYB,BSQA,ERFA
     1Y,ERFBY,FACTN1,FACTN2,SCALE1,SCALE2,TWON,U,V,W,WW,DELTAY
C
      DIMENSION F(101),G(101)
C
C    SET UPPER LIMIT,Y INCREMENT,INITIAL VALUES,AND CALCULATE CONSTANTS
C
      O=10.0**(-77)
      P=10.0**73
      SCALE1=0.1D-66
      SCALE2=0.1D-68
      ULIMIT=0.8D0
      LIMITU=25
      DELTAY=0.1D0
      WRITE (12,202) O,P,SCALE1,SCALE2,LIMITU,DELTAY
C
      R=0.9D0
      A=1.D0/(2.D0*(1.D0-R*R))
```

```
      SQA=DSQRT(A)
      CB=(-1.D0*R)/(1.D0-R*R)
      CC=1.D0/(2.D0*(1.D0-R*R))
      D=1.D0/(DSQRT(3.141592653589)*0.436D0*SQA)
      E=2.D0/DSQRT(3.141592653589)
      G(1)=0.0
      Y=0.D0
C
C     DO LOOP TO COMPUTE VOLUME UNDER THE P.D.F. USING THE TRAPEZOIDAL RULE
C
      DO 2 J=1,LIMITU
      B=CB*Y
      C=CC*Y*Y
      ABC=(B*B-A*C)/A
      AFYB=SQA*0.246D0*DSQRT(3.05D0+Y*Y)+B/SQA
      BSQA=B/SQA
C
C     DO LOOP TO COMPUTE REQUIRED ERROR FUNCTIONS BY SERIES EXPANSION
C

      WRITE (12,200)
      N=0
      S=0.D0
      FACTN1=1.D0*SCALE1
      FACTN2=1.D0*SCALE2
      TWON=1.D0
      ERFAY=AFYB
      ERFBY=BSQA
C
      DO 1 I=1,100
C
      N=N+1
      NN=N+1
      S=S+1.D0
C
```

```
C     COMPUTATION OF N FACTORIAL FOR USE IN ERROR FUNCTION CALCULATION
C
      FACTN1=S*FACTN1
      FACTN2=S*FACTN2
      TWON=TWON+2.D0
C
      U=((AFYB**N)*SCALE1)*(AFYB**NN)
      V=((BSQA**N)*SCALE2)*(BSQA**NN)
      W=FACTN1*TWON
      WW=FACTN2*TWON
      UMAG=DABS(U)
      VMAG=DABS(V)
C
C     ARITHMETIC IF STATEMENTS TO STOP SUMMATION OF ERROR FUNCTION TERMS WHEN
C     EITHER 0.1E-74 OR 0.1E 75 IS EXCEEDED. IF IN 125 REPETITIONS OF THE SUMMA-
C     TION NEITHER OF THESE TWO LIMITS ARE REACHED, THE SUM IS TERMINATED.
C
      IF (J-10)5,5,6
    5 IF (O-UMAG)7,7,3
    7 IF (O-VMAG)6,6,3
    6 IF (P-W)3,3,9
    9 IF (P-WW)3,3,8
    8 CONTINUE
C
C     COMPUTATION OF ERROR FUNCTIONS ERFAY AND ERFBY
C
      ERFAY=((-1.D0)**N)*U/W+ERFAY
      ERFBY=((-1.D0)**N)*V/WW+ERFBY
      WRITE (12,201) ERFAY,ERFBY,N,U,W,V,WW
    1 CONTINUE
C
    3 ERFAY=E*ERFAY
      ERFBY=E*ERFBY
      F(J)=DEXP(ABC)*(ERFAY-ERFBY)
```

```
      WRITE (12,205)
      WRITE (12,206) Y,F(J),AFYB,ERFAY,BSQA,ERFBY
      Y=Y+DELTAY
C
C    ARITHMETIC IF STATEMENT ALLOWS TERM F(1) TO BE COMPUTED PRIOR TO APPLICATION

C    OF THE TRAPEZOIDAL RULE.
C
      IF(J-1)2,2,4
    4 G(J)=G(J-1)+DELTAY*0.5*(F(J)+F(J-1))
    2 CONTINUE
      BETA=1.00-G(LIMITU)*D
      WRITE (12,207) BETA
      WRITE (12,208) (G(J),J=1,LIMITU)
      END
```